SeqWare version 4.0
Copyright (c) 1993 - 2000 Compugen Ltd.

SM search: mode: search, minimum sw score:

Run on: November 23, 2001, 12:08:42 ; Search time 28 seconds
(without alignments)
41006.429 Million cell updates/sec

Title: US-09-580-515-1
Perfect score: 1323
Sequence: 1 atgaaaaaatatttatacc.....aatccgat ttcatcgctaa 1323

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 951203 seqs, 1132380999 residues

Total number of hits satisfying chosen parameters: 951203

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/backup/5_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/backup/6_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/backup/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backup/Terminated.seq:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1322	99.9	1323	4	US-09-259-214-1	Sequence 1, Appli
2	1322	99.9	1323	4	US-09-318-528-1	Sequence 1, Appli
3	1322	99.9	1323	4	US-09-291-931-1	Sequence 1, Appli
4	1213	91.7	1212	2	US-08-910-798-1	Sequence 1, Appli
5	41	3.1	908	2	US-07-885-898A-5	Sequence 5, Appli
6	33	2.5	2511	2	US-08-680-329-116	Sequence 116, App
7	33	2.5	45100	2	US-08-770-479-17	Sequence 17, Appl
8	33	2.5	45100	4	US-08-757-669A-17	Sequence 17, Appl
9	32.8	2.4	44	4	US-09-103-840A-2	Sequence 2, Appli
10	31.6	2.4	8160	1	US-08-469-005A-9	Sequence 9, Appli
11	31.6	2.4	8519	1	US-09-261-907-1	Sequence 1, Appli
12	31	2.3	4142	1	US-08-441-430-3	Sequence 3, Appli
13	31	2.3	4488	1	US-08-441-430-1	Sequence 1, Appli
14	30.4	2.3	8252	1	US-08-046-585-15	Sequence 15, Appl
15	30.4	2.3	8252	1	US-08-493-763-15	Sequence 15, Appl
16	30.4	2.3	8252	5	PCT-US93-11721-15	Sequence 15, Appl
17	29.6	2.2	150	4	US-08-659-473-8	Sequence 8, Appli
18	29.6	2.2	1005	2	US-08-875-062-4	Sequence 4, Appli
19	29.6	2.2	15211	1	US-08-496-855A-3	Sequence 3, Appli
20	29.6	2.2	15211	2	US-07-938-154-9	Sequence 9, Appli
21	29.6	2.2	15211	5	PCT-US91-02311-9	Sequence 9, Appli
22	29.6	2.2	24??	2	US-08-466-589-9	Sequence 9, Appli
23	29.6	2.2	2450	2	US-08-708-646-9	Sequence 9, Appli
24	29.6	2.2	2450	3	US-08-467-574-9	Sequence 9, Appli
25	29.6	2.2	2450	4	US-07-217-345-9	Sequence 9, Appli
26	29.4	2.2	667	1	US-08-499-215-1	Sequence 1, Appli
27	29.2	2.2	508	3	US-08-669-408B-3	Sequence 3, Appli
28	29.2	2.2	1555	3	US-08-669-408B-9	Sequence 9, Appli
29	29	2.2	1548	2	US-08-762-106-5	Sequence 5, Appli
30	29	2.2	1548	4	US-09-320-774-5	Sequence 5, Appli
31	29	2.2	1581	2	US-08-762-106-6	Sequence 6, Appli
32	29	2.2	1581	4	US-09-320-774-6	Sequence 6, Appli
33	29	2.2	3531	1	US-08-629-606-1	Sequence 1, Appli
34	29	2.2	4531	4	US-09-076-761-1	Sequence 1, Appli
35	29	2.2	56516	2	US-08-906-306-1	Sequence 1, Appli
36	29	2.2	56516	4	US-09-338-907-1	Sequence 1, Appli
37	29	2.2	56520	4	US-09-338-907-179	Sequence 179, App
38	28.8	2.2	4411529	4	US-09-103-840A-1	Sequence 1, Appli
39	28.6	2.2	1897	1	US-08-453-477-1	Sequence 1, Appl
40	28.6	2.2	1897	1	US-08-453-461-1	Sequence 1, Appl
41	28.4	2.1	243	1	US-08-741-950B-26	Sequence 26, Appl
42	28.4	2.1	243	4	US-09-083-756A-26	Sequence 26, Appl
43	28.4	2.1	519	3	US-08-860-165-13	Sequence 13, Appl
44	28.4	2.1	519	4	US-09-359-382-13	Sequence 13, Appl
45	28.4	2.1	840	1	US-08-434-255-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-259-214-1
; Sequence 1, Application US/09259214A
; Patent No. 6116770
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 10459-47A-1
; CURRENT APPLICATION NUMBER: US/09/259,214A
; CURRENT FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 60/076,798
; PRIOR FILING DATE: 1998-03-04
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-09-259-214-1

Query Match 99.9%; Score 1322; DB 3; Length 1323;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaaaaaatatttatacctattttatcttctgattcagttaaccccgaatctgtcgca 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 atgaaaaaatatttatacctattttatcttctgattcagttaaccccgaatctgtcgca 60

Qy 61 ttgcctgctaatgaagccgaagctgaagctgaaaagttgcagttgattgatcaagtcatt 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 ttgcctgctaatgaagccgaagctgaagctgaaaagttgcagttgattgatcaagtcatt 120

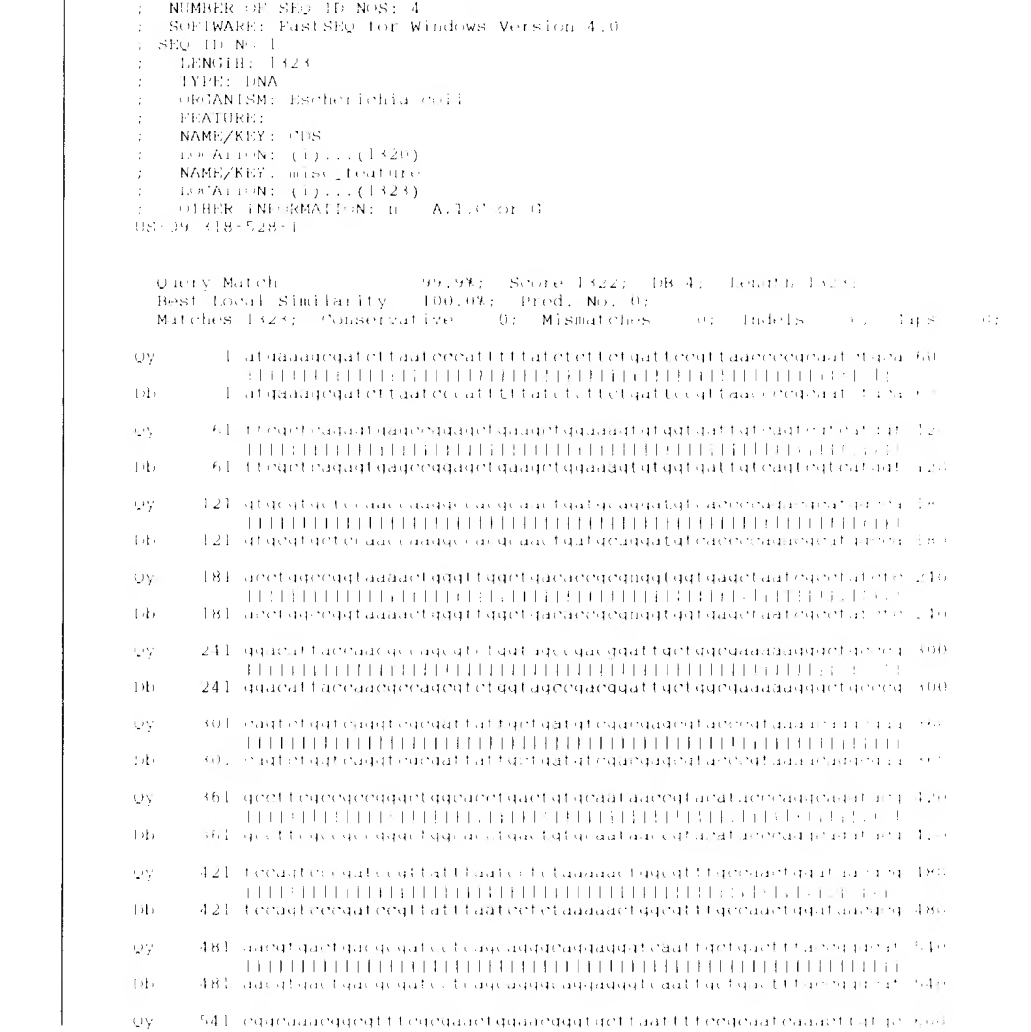
Qy 121 gtgcctgctgctgtcaaccaaagccaagcaactgatgccgaatgtcgccccaaacgctgccg 180
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 gtgcctgctgctgtcaaccaaagccaagcaactgatgccgaatgtcgccccaaacgctgccg 180

Qy 181 acctgccgcgatgaaactgagttgcctgacaccgccgtgattgatgaactgatgcctatcc 240
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 acctgccgcgatgaaactgagttgcctgacaccgccgtgattgatgaactgatgcctatcc 240

Qy 241 ggacattaccgaacgccagcgtctgattgaccgaacgattgctgtgccgaaaaagatttcc 300
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 ggacattaccgaacgccagcgtctgattgaccgaacgattgctgtgccgaaaaagatttcc 300

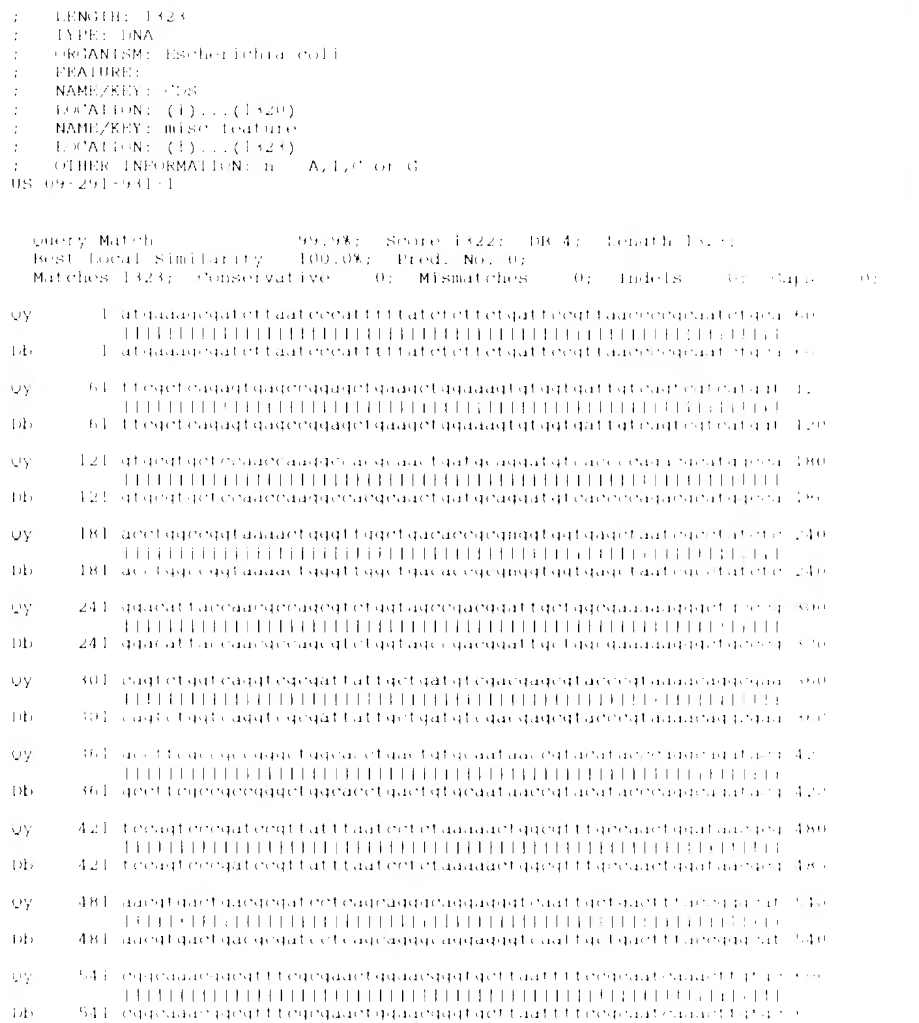
RESULT 2
US-09-318-528-1
; Sequence 1, Application us/09318528
; Patent No. 6184340
; GENERAL INFORMATION:
; APPLICANT: Krotz, Keith
; TITLE OF INVENTION: Novel Promoter
; FILE REFERENCE: 09810/029003
; CURRENT APPLICATION NUMBER: US/09/318,528
; CURRENT FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 60/291,931
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 07/259,214
; EARLIER FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-09-318-528-1

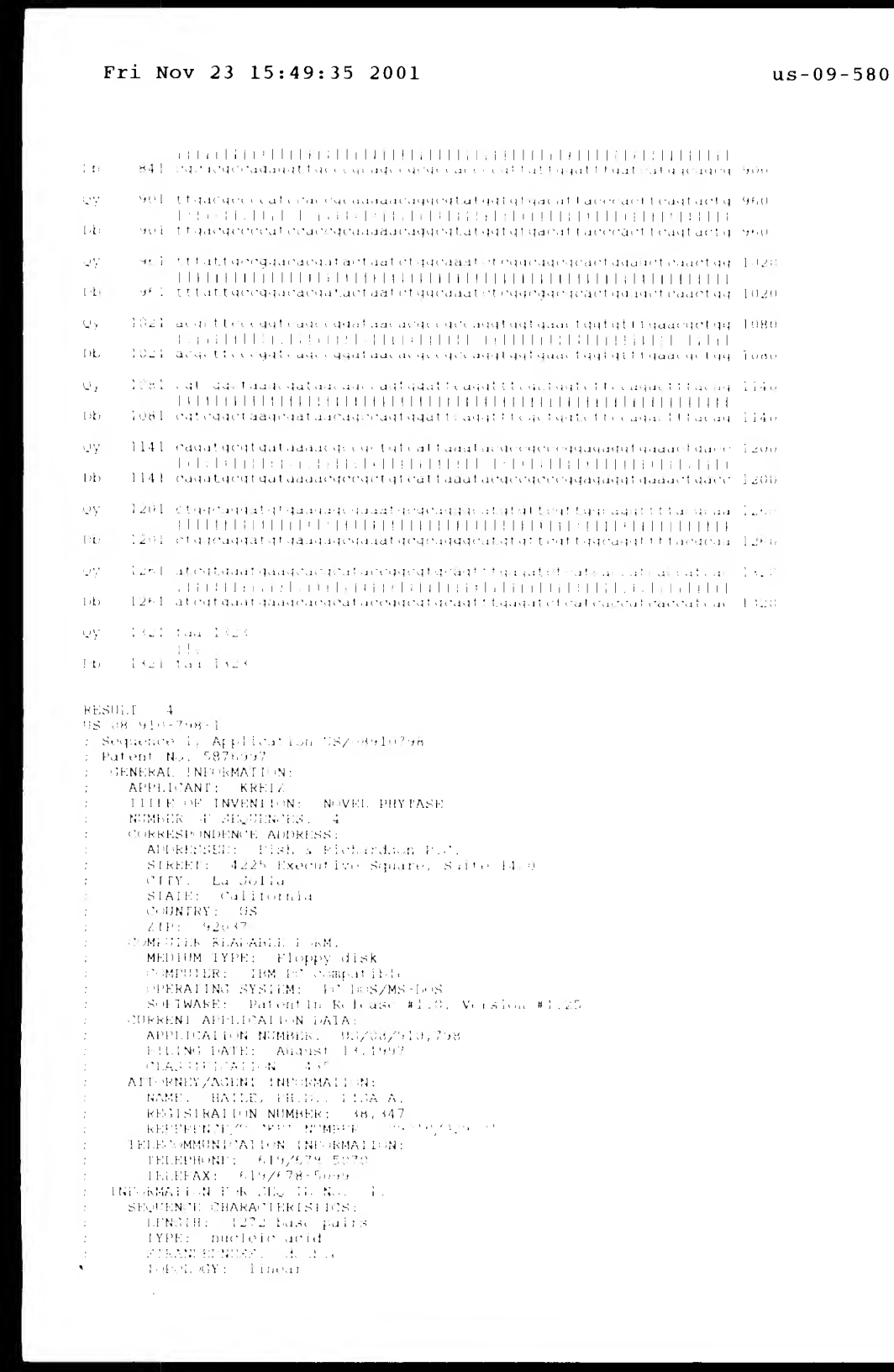
Query Match 99.9%; Score 1322; DB 4; Length 1323;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



RESULT 4
US-09-291-931-1
; Sequence 1, Application US/09/291931A
; Patent No. 6190907
; GENERAL INFORMATION:
; APPLICANT: Kretz, Kevin
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 5001/0/1/US
; CURRENT APPLICATION NUMBER: US/09/291,931A
; CURRENT FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 09/250,214
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 09/135,758
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ IDS: 4
; SOFTWARE: PatentIn for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-09-291-931-1

Query Match 99.9%; Score 1322; DB 4; Length 1323;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





RESULT 4
US-08-910-798-1
; Sequence 1, Application US/08910798
; Patent No. 5876997
; GENERAL INFORMATION:
; APPLICANT: KRETZ
; TITLE OF INVENTION: NOVEL PHYTASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,798
; FILING DATE: August 13,1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BATTE, RICHARD A.
; REGISTRATION NUMBER: 38,847
; REFERENCE/DOCKET NUMBER: 07039/035001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: PHYTASE
; FEATURE:
; NAME/KEY:
; LOCATION:
US-08-910-798-1

Query Match 91.5%; Score 1210; DB 2; Length 1272;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 1272; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

```
Qy      901 tttaaggcccaatcccaccccaaaaaacaaggcgtatggtgcgtaatttacccacttcaatgactcg  960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      850 TTTAGGGCCCAATCCACCCCAAAAACACAGGCGTATGGTGCGTAATTTACCCACTTCAATGACTCG  909

Qy      961 ttttatttccaqcaacaqatactaatctqqcaatctcqqcqqcqcqctqqaqctcaactcqq 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      910 TTTATTTCCTGCACAGATACTAATCTGCCAAATCTCGGCGGCGCGCTGGAGCTCAACTCGG 969

Qy      1021 ctcqqctcaaqqcqqcqqaqqqcqqcqqcqqaqqtqqtqqaqqtqqaqqtqqtttqaqqcqqtcq 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      970 CTCGGCTCAAGGCGGCGGAGGGCGGCGGCGGAGGTGGTGGAGGTGGAGGTGGTTTGAGGCGGTCG 1029

Qy      1081 cqtqqctcaaqtqqcqqcqqqqtqqtqqtqqtqqtqqtttqqtqqtqqtqqtttqqtqqcqtcaq 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1030 CGTGGCTCAAGTGGCGGCGGGGTGGTGGTGGTGGTGGTTTGGTGGTGGTGGTTTGGTGGCGTCAG 1089

Qy      1141 cqqatqcqtqatqaaaqcqcqqctqtcaatqaatqcqcqcqcqcqaaqqtqaaqcqaacqtqcqcc 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1090 CAGATGCGTGATAAAGCGCGGCTGTCATGAATGCGCGCGCGCGAAGGTGAAGCGAACGTGCGCC 1149

Qy      1201 ctqqccqaaqtqatqaaqcqcqaatqcqcqqqqcqcqqqqtqatqqtcatqqcqaqtqaaqcqaac 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1150 CTGCCGAGGATGATGAAGCGCGAATGCGCGGGCGCGGGGTGATGGTCATGGCGAGTGAAGCGAAC 1209

Qy      1261 atcqtqaatqaaqcqaqcqatacccqtqtcqqttqaatqtatcaccqatacccaatacqc 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1210 ATCGTGAATGAAGCGAGCGATACCCGTGTCGGTTGAATGTATCACCGATACCCAATACGC 1269

Qy      1321 taa 1323
             |||
Db      1270 TAA 1272


RESULT  5
US-07-895-080B-5/2
; Sequence 5, Application US/07895080B
; Patent No. 5830705
; GENERAL INFORMATION:
;     APPLICANT:  Shoyab, Mohammed
;     APPLICANT:  McDonald, Vicki L.
;     APPLICANT:  Bradley, James G.
;     APPLICANT:  Plowman, Gregory D.
;     TITLE OF INVENTION:  AMPHIREGULINS:  A FAMILY OF
;     TITLE OF INVENTION:  HEPARIN BINDING EPITHELIAL CELL GROWTH FACTORS
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fenwick & Edmonds LLP
;       STREET:  1185 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC Compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  07/895,080B
;       FILING DATE:  18 MAY 1992
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Cohen, Laura A.
;       REGISTRATION NUMBER:  32,212
;       REFERENCE/DOCKET NUMBER:  5624 174
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212 790 9090
;       TELEFAX:  212 869 9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  909 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..744
US-07-895-080B-5


  Query Match              2.6%;  Score 34;  DB 2;  Length 909;
  Best Local Similarity   56.1%;  Pred. No. 0.44;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    3;  Gaps    1;

Qy       4 aaagcqatcttaatccaatttttatctcttctqattcqttaaccccqaatctqcatctc 63
           ||||  |||||| ||  || ||  || ||  |  |||||  |  |||||||||||||||
Db     510 AAAGTAATCTTGATGACAATTGCATGTCACCACCTCCAGGTTCTCGATGTATCTCGCATTC 451

Qy      64 gctcaqatqaaqccqaaqctqaaqctqaaaqtqatqtqatqattatcaqtcatcat 117
           ||   || ||    |  |||| || |  ||||    |||| |  || ||||||||
Db     450 CGCATGAATCCAAAAGTTCTGAAACTTCGCAGTGCATCGATTCTTTTCTTCTT 397


RESULT  6
US-08-680-326-116/C
; Sequence 116, Application US/08680326
; Patent No. 5925733
; GENERAL INFORMATION:
;     APPLICANT:  ROSE, TIMOTHY M.
;     APPLICANT:  BOSCH, MARNIX
;     APPLICANT:  STRAND, KURT
;     APPLICANT:  DOLARO, GEORGE J.
;     TITLE OF INVENTION:  DNA POLYMERASE OF GAMMA HERPES VIRUSES
;     TITLE OF INVENTION:  ASSOCIATED WITH KAPOSI'S SARCOMA AND LYMPHOTROPIC NUCL
;     TITLE OF INVENTION:  FIBROMATOSIS
;     NUMBER OF SEQUENCES:  152
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORRISON & FOERSTER
;       STREET:  755 Page Mill Road
;       CITY:  Palo Alto
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94304-1018
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  08/680,326
;       FILING DATE:
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Schiff, J. Michael
;       REGISTRATION NUMBER:  40,253
;       REFERENCE/DOCKET NUMBER:  29938.20001.00
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 813-5600
;       TELEFAX:  (415) 494-0792
;       TELEX:  706141
;   INFORMATION FOR SEQ ID NO:  116:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2511 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..2511
US-08-680-326-116


  Query Match              2.5%;  Score 33;  DB 2;  Length 2511;
  Best Local Similarity   40.7%;  Pred. No. 1.1;
```

Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

RESULT 7
US-08-770-379-17/c
; Sequence 17, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Botchesky, Roy A.
; APPLICANT: Bossi, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: [illegible]
; TITLE OF INVENTION: [illegible]
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-770-379-17

Query Match 2.5%; Score 33; DB 2; Length 35100;
Best Local Similarity 49.7%; Pred. No. 4.8;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

RESULT 8
US-08-757-669A-17/c
; Sequence 17, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Botchesky, Roy A.
; APPLICANT: Bossi, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: <5100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-757-669A-17

Query Match 2.5%; Score 33; DB 4; Length 35100;
Best Local Similarity 49.7%; Pred. No. 4.8;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

RESULT 9
US-09-165-646A-2/c
; Sequence 2, Application US/09165646A
; Patent No. 6294328
; GENERAL INFORMATION:

```
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366.26007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 44-3765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;    OTHER INFORMATION: CDC 1551
;    OTHER INFORMATION: "n" bases at various positions throughout the sequence
;    OTHER INFORMATION: represent a, t, g or c
US-09-103-840A-2

  Query Match             2.1%;  Score 31.8;  DB 1;  Length 110765;
  Best Local Similarity   47.4%;  Pred. No. 46;
  Matches   93;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy      196 ctgggttggctgacacc[illegible]gcggtggtgagctaatcgcctatctcggacattaccaacgc 255
              |||  |  |  ||    | |    |||| ||   |      ||        ||  ||    ||||
Db   926394 CTGGCGGTGTGCAGCGGGTTGGTGCTGGCCGGGGCGTTGGCGCCGTCCAAACCCAACAA 926335

Qy      256 [illegible]ctgqtagccgacggattgctggcgaaaaaagctgcccgcagtctggtcagatc 315
              ||||      |   ||||  ||           ||          ||||  |  |  ||  ||  |       |  |
Db   926334 CAGCCCACCGATGCCGCCGACGCCGGTGCCGCCGGGCATCGCCCATTCCGCCACTGCC 926275

Qy      316 gcgatttattgctgatgtcgacgagcgtacccgtaaaacagacgaagccttcgccgccggg 375
              ||   ||     ||  |  || |||   |||        ||            |  ||||   |||   |  |||||
Db   926274 [illegible]ACCGCCGTT 926215

Qy      376 ctggcactgactgtg 391
              |    |    ||      |  ||
Db   926214 [illegible]GGTG 926199


RESULT 10
US-08-469-005A-9/c
; Sequence 9, Application US/08469005A
; Patent No. 5977474
;  GENERAL INFORMATION:
;     APPLICANT:  KGHAJDA, FRANCIS P.
;     APPLICANT:  PASTERNACK, GARY A.
;     TITLE OF INVENTION:  CANCER RELATED ANTIGEN
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  BAKER & BOTTS, L.L.P.
;        STREET:  1299 Pennsylvania Avenue, N.W.
;        CITY:  Washington
;        STATE:  DC
;        COUNTRY:  USA
;        ZIP:  20004-2400
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE:  Diskette
;        COMPUTER:  IBM Compatible
;        OPERATING SYSTEM:  DOS
;        SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER:  US/08/469,005A
;        FILING DATE:  06-JUN-1995
;        CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;        APPLICATION NUMBER:  08/186,129
;        FILING DATE:  24-JAN-1994
;        APPLICATION NUMBER:  08/090,908
;        FILING DATE:  26-JUL-1993
;        APPLICATION NUMBER:  07/917,716
;        FILING DATE:  24-JUL-1992
;     ATTORNEY/AGENT INFORMATION:
;        NAME:  Posorske, Laurence H
;        REGISTRATION NUMBER:  34,698
;        REFERENCE/DOCKET NUMBER:  062482-0113
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE:  202-639-7700
;        TELEFAX:  202-639-7890
;        TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH:  8460 base pairs
;        TYPE:  nucleic acid
;        STRANDEDNESS:  single
;        TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:
;     FEATURE:
;        NAME/KEY:  Coding Sequence
;        LOCATION:  124...7650
;        OTHER INFORMATION:
US-08-469-005A-9

  Query Match             2.4%;  Score 31.6;  DB 1;  Length 8460;
  Best Local Similarity   47.9%;  Pred. No. 6.4;
  Matches   91;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy      712 atgctgacggagatatttctcctgcaacaagcacagggaatgcc[illegible] 771
              ||||||| |   |    |       | ||  |    |      |  ||| |   | |     | |
Db     7604 ATGCTGATGATGGACTCCAGGCCGCTGCCCTCCAGCAGCGTGCGGTGGTCACCCTCGATC 7545

Qy      772 aggatcaccgattcacaccagtggaacaccttgctaagtttgcataacgcgcatttt[illegible] 831
              | |     || ||| |      || ||  |  |  ||||  |      ||  |||  |     |||||         |
Db     7544 ACGTGGACGGATACTTTCCGTCGCATACCTGGGAGAGGTTGTAATC[illegible] 7485

Qy      832 ttgctacaacgcacgccagaggttgcccgcagccgcgccacccgttattggatttgatc 891
              | ||       |        | || |      | ||  |||||    ||    ||       |    || |  |        |
Db     7484 [illegible]AGTAGCATCACGTTGCCATCGTATT[illegible] 7425

Qy      892 atggcagcgt 901
               | |    |  ||
Db     7424 TTGGGTGTGT 7415


RESULT 11
US-09-261-907-1/c
; Sequence 1, Application US/09261907A
; Patent No. 6294364
; GENERAL INFORMATION:
;  APPLICANT: ELLIS, CATHERINE
;  APPLICANT: LONSDALE, JOHN
;  APPLICANT: BERGSMA, DERK J.
;  APPLICANT: MOONEY, JEFFREY L.
;  APPLICANT: DePIERA, MEGAN E.
;  APPLICANT: CHAPMAN, CONRAD
;  TITLE OF INVENTION: HUMAN FAS
;  FILE REFERENCE: GP-70603
;  CURRENT APPLICATION NUMBER: US/09/261,907A
;  CURRENT FILING DATE: 1999-03-03
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 8519
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-261-907-1
```

```
  Query Match              2.4%;  Score 31.6;  DB 4;  Length 8519;
  Best Local Similarity   47.2%;  Pred. No. 5.4;
  Matches   91;  Conservative    0;  Mismatches   99;  Indels    9;  Gaps    3;

Qy      712 atatctaaaaaaaaaatattttctcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 771
            ||||||| | |   |     | || |   |     |  ||| | | |    | |
Db     7542 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 7601

Qy      772 aaaatcaccaattcaccaaaatgaaaccacctgctaaatttgcatcaaaccaaattttat 831
            ||  ||  |||||  |  |||| | | | | |  | |||  |  ||||   |     |
Db     7602 AAGTGGAACGGATATTTTCCGTTCGATACTTCGGAGAGGTTGTAGTCGCGCGCCAGGTCC 7523

Qy      832 ttactaaaaaaaaaaaaaaaatttgccaaaaccaccaaaccccattaaaatttgatc 891
            |||  |    |  | | |  | ||| |||  ||  |||  |  | | | | |||  |
Db     7522 TCGCGCTAGGCGCCACCCGTCTTGGCGCGCAGTAGCCATCACGTTGCCATGCTACTTGGCC 7463

Qy      892 atagcaacgt 901
            |||  | ||
Db     7462 TTGGGTGTGT 7453


RESULT  12
US-08-441-430-4
; Sequence 4, Application US/08441430
; Patent No. 5681942
;  GENERAL INFORMATION:
;    APPLICANT:  Buchwald, Manuel
;    APPLICANT:  Strathdee, Craig A.
;    APPLICANT:  Wevrick, Rachel
;    APPLICANT:  Mathew, Christopher George Porter
;    TITLE OF INVENTION:  Fanconi Anemia Type C Gene
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Polley, Esq.
;      ADDRESSEE:  Klarquist, Sparkman, Campbell, Leigh &
;      ADDRESSEE:  Whinston, LLP
;      STREET:  121 S.W. Salmon, Suite 1600
;      CITY:  Portland
;      STATE:  Oregon
;      COUNTRY:  U.S.A.
;      ZIP:  97204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Disk, 5+ inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  WordPerfect 5.1/5.2 ASCII Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,430
;      FILING DATE:  May 15, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 07/876,285
;      FILING DATE:  April 29, 1992
;      APPLICATION NUMBER:  U.S. 07/918,413
;      FILING DATE:  July 21, 1992
;      APPLICATION NUMBER:  U.S. 08/003,963
;      FILING DATE:  January 15, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Richard J. Polley, Esq.
;      REGISTRATION NUMBER:  28,107
;      REFERENCE/DOCKET NUMBER:  3812-42824
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (503) 226-7391
;      TELEFAX:  (503) 228-9446
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4147 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Double stranded
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      IMMEDIATE SOURCE:
;        LIBRARY:  Human cDNA
;      POSITION IN GENOME:  (of corresponding genomic gene)
;        CHROMOSOME/SEGMENT:  9q
;        MAP POSITION:  22.3
;      UNITS:
US-08-441-430-4


  Query Match              2.4%;  Score 31;  DB 1;  Length 9447;
  Best Local Similarity   57.9%;  Pred. No. 5.8;
  Matches   55;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy      601 ctttaaacgtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 660
            || | | | | | ||||  ||| | |   || | ||   |||||||  | |||| |
Db     2953 CTCCAGCCTGGGCAACAGAGCGAGACCATCTCTTAAAAAAAAGGCATTGTTAGTTAACTC 2912

Qy      661 aaaatgaaaccccaacaatatatctcattaaccaatgc 695
            ||||| | |    |     | ||| ||   ||||
Db     2911 TCTCAAGTTAATAATTATTATTTTTTCATTAAAAATGC 2877


RESULT  13
US-08-441-430-1
; Sequence 1, Application US/08441430
; Patent No. 5681942
;  GENERAL INFORMATION:
;    APPLICANT:  Buchwald, Manuel
;    APPLICANT:  Strathdee, Craig A.
;    APPLICANT:  Wevrick, Rachel
;    APPLICANT:  Mathew, Christopher George Porter
;    TITLE OF INVENTION:  Fanconi Anemia Type C Gene
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Polley, Esq.
;      ADDRESSEE:  Klarquist, Sparkman, Campbell, Leigh &
;      ADDRESSEE:  Whinston, LLP
;      STREET:  121 S.W. Salmon, Suite 1600
;      CITY:  Portland
;      STATE:  Oregon
;      COUNTRY:  U.S.A.
;      ZIP:  97204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Disk, 5+ inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  WordPerfect 5.1/ASCII Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,430
;      FILING DATE:  May 15, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 07/876,285
;      FILING DATE:  April 29, 1992
;      APPLICATION NUMBER:  U.S. 07/918,413
;      FILING DATE:  July 21, 1992
;      APPLICATION NUMBER:  U.S. 08/003,963
;      FILING DATE:  January 15, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Richard J. Polley, Esq.
;      REGISTRATION NUMBER:  28,107
;      REFERENCE/DOCKET NUMBER:  3812-42824
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (503) 226-7391
;      TELEFAX:  (503) 228-9446
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4488 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Double stranded
;      TOPOLOGY:  Linear
```

```
;    MOLECULE TYPE: cDNA to mRNA
;    HYPOTHETICAL: NO
;    ANTI-SENSE: NO
;    ORIGINAL SOURCE:
;      ORGANISM: Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY: Human cDNA
;    POSITION IN GENOME: (of corresponding genomic gene)
;      CHROMOSOME/SEGMENT: 9q
;      MAP POSITION: 22.3
;      UNITS:
US-08-441-130-1

  Query Match            2.3%;  Score 31;  DB 1;  Length 1108;
  Best Local Similarity  57.9%;  Pred. No. 7.1;
  Matches   55;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     601 [illegible] 660

Db    2871 [illegible]

Qy     661 [illegible] 695

Db    2931 AAGTTAACATTATTCATGTCAGTACAGGTGT 2965


RESULT 11
US-08-047-585-15
; Sequence 15, Application US/08047585
; Patent No. 5455901
; GENERAL INFORMATION:
;    APPLICANT: Lamarco, Kelly
;    APPLICANT: Wilson, Angus
;    APPLICANT: Herr, Winship
;    TITLE OF INVENTION: A NOVEL EUKARYOTIC TRANSCRIPTION PROTEIN:
;    TITLE OF INVENTION: HOST CELL FACTOR
;    NUMBER OF SEQUENCES: 15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET: 4 Embarcadero Center, Suite 3400
;      CITY: San Francisco
;      STATE: CA
;      COUNTRY: USA
;      ZIP: 94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/047,585
;      FILING DATE: 12-APR-1993
;      CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Osman, Richard A
;      REGISTRATION NUMBER: 36,627
;      REFERENCE/DOCKET NUMBER: A-57593-1/RA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (415) 781-1989
;      TELEFAX: (415) 398-3249
;      TELEX: 910-277299
; INFORMATION FOR SEQ ID NO: 15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 8252 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;    MOLECULE TYPE: cDNA
US-08-047-585-15

  Query Match            2.3%;  Score 30.4;  DB 1;  Length 8252;
  Best Local Similarity  51.5%;  Pred. No. 16;
  Matches   70;  Conservative    0;  Mismatches   66;  Indels   [illegible];  Gaps   [illegible];

Qy      47 [illegible]

Db    4681 [illegible]

Qy     107 [illegible]

Db    4741 [illegible]

Qy     167 [illegible] 182

Db    4801 [illegible] 4816


RESULT 12
US-08-393-703-15
; Sequence 15, Application US/08393703
; Patent No. 5851239
; GENERAL INFORMATION:
;    APPLICANT: Lamarco, Kelly
;    APPLICANT: Wilson, Angus
;    APPLICANT: Herr, Winship
;    TITLE OF INVENTION: A NOVEL EUKARYOTIC TRANSCRIPTION PROTEIN
;    TITLE OF INVENTION: HOST CELL FACTOR
;    NUMBER OF SEQUENCES: 15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET: 4 Embarcadero Center, Suite 3400
;      CITY: San Francisco
;      STATE: CA
;      COUNTRY: USA
;      ZIP: 94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/393,703
;      FILING DATE: 24-FEB-1995
;      CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Osman, Richard A
;      REGISTRATION NUMBER: 36,627
;      REFERENCE/DOCKET NUMBER: A-57593-2/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (415) 781-1989
;      TELEFAX: (415) 398-3249
;      TELEX: 910-277299
; INFORMATION FOR SEQ ID NO: 15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 8252 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;    MOLECULE TYPE: cDNA
US-08-393-703-15

  Query Match            2.3%;  Score 30.4;  DB 1;  Length 8252;
  Best Local Similarity  51.5%;  Pred. No. 16;
  Matches   70;  Conservative    0;  Mismatches   66;  Indels   [illegible];  Gaps   [illegible];

Qy      47 [illegible]

Db    4681 [illegible]

Qy     107 [illegible]

Db    4741 [illegible]
```

qy 167 ccaaggatgtccaaa 182
11 4861 [illegible] 4816

Search completed: November 22, 2001, 17:00:58
Job time: 1154+ sec

GenCore version 4.5
Copyright (c) 1993 - 2001 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2001, 09:58:43 ; Search time 2.47 Seconds
(without alignments)
78.412 Million cell updates/sec

Title: US-09-580-515-2
Perfect score: 2302
Sequence: 1 MKATLIFFLSLLIPLIPQSA.....IVNEAKIPAVSLRSHHHHHH 440

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 21252 seqs, 22503292 residues

Total number of hits satisfying chosen parameter: 21252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/backfile1.pep:*
6: /cgn2_6/ptodata/1/iaa/backfile2.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB	ID	Description
1	2302	100.0	440	5	US-09-259-214-2	Sequence 2, Appli
2	2302	100.0	440	1	US-09-318-528-2	Sequence 2, Appli
3	2302	100.0	440	4	US-09-291-981-2	Sequence 2, Appli
4	2194.5	95.3	439	2	US-08-940-798-2	Sequence 2, Appli
5	119	5.2	167	4	US-08-868-435-33	Sequence 33, Appl
6	117	5.1	386	1	US-08-758-214-1	Sequence 1, Appli
7	117	5.1	386	2	US-08-940-798-1	Sequence 33, Appl
8	117	5.1	669	4	US-09-097-199-48	Sequence 48, Appl
9	117	5.1	945	2	US-09-146-203-2	Sequence 2, Appli
10	117	5.1	945	4	US-08-979-824A-2	Sequence 2, Appli
11	117	5.1	945	4	US-09-344-196-2	Sequence 2, Appli
12	116	5.0	965	5	PCT-US95-04610-11	Sequence 14, Appl
13	101.5	4.4	479	1	US-07-655-721-2	Sequence 2, Appli
14	101.5	4.4	479	2	US-08-609-126A-2	Sequence 2, Appli
15	101.5	4.4	479	5	US-08-474-055-4	Sequence 4, Appl
16	99.5	4.3	405	4	US-08-808-435-12	Sequence 12, Appl
17	97	4.2	443	4	US-08-963-459-28	Sequence 28, Appl
18	93.5	4.1	1150	4	US-09-428-517-2	Sequence 2, Appli
19	93	4.0	2184	5	US-08-246-113-5	Sequence 5, Appli
20	92	4.0	438	5	US-08-080-506-3	Sequence 3, Appli
21	92	4.0	439	5	US-08-080-506-7	Sequence 7, Appli
22	92	4.0	3033	1	US-07-935-696-5	Sequence 5, Appli
23	91.5	4.0	466	4	US-08-868-435-2	Sequence 2, Appli
24	91.5	4.0	645	1	US-08-184-427A-4	Sequence 4, Appli
25	91.5	4.0	645	2	US-08-078-311-1	Sequence 1, Appli
26	91.5	4.0	645	2	US-08-460-482-1	Sequence 1, Appli
27	91.5	4.0	645	5	PCT-US95-00676-4	Sequence 4, Appli
28	90.5	3.9	413	4	US-08-963-459-30	Sequence 30, Appl
29	90.5	3.9	1085	1	US-08-485-588-5	Sequence 5, Appli
30	90.5	3.9	1085	1	US-08-484-565-5	Sequence 5, Appli
31	90.5	3.9	1085	2	US-08-480-751-5	Sequence 5, Appli
32	90.5	3.9	1085	2	US-08-944-986-5	Sequence 5, Appli
33	90.5	3.9	1085	4	US-08-854-784-5	Sequence 5, Appli
34	90.5	3.9	1085	4	US-08-484-7198-5	Sequence 5, Appli
35	90	3.9	626	1	US-08-164-427A-2	Sequence 2, Appli
36	90	3.9	626	5	PCT-US95-00670-2	Sequence 2, Appli
37	90	3.9	921	1	US-08-188-582-2	Sequence 2, Appli
38	90	3.9	921	1	US-08-646-715-2	Sequence 2, Appli
39	89.5	3.9	598	4	US-09-352-159-42	Sequence 42, Appl
40	89	3.9	633	1	US-08-250-859-17	Sequence 17, Appl
41	89	3.9	633	1	US-08-490-803-17	Sequence 17, Appl
42	89	3.9	633	2	US-08-457-254-7	Sequence 7, Appli
43	89	3.9	633	5	PCT-US94-08806-17	Sequence 17, Appl
44	89	3.9	633	5	PCT-US95-01225-17	Sequence 17, Appl
45	89	3.9	633	5	PCT-US95-10626-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-259-214-2
; Sequence 2, Application US/09259214A
; Patent No. 6110719
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: PF0581/70-1
; CURRENT APPLICATION NUMBER: US/09/259,214A
; CURRENT FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US/08/940,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-259-214-2

Query Match 100.0%; Score 2302; DB 5; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.4e-234;
Matches 440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKATLIFFLSLLIPLIPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MKATLIFFLSLLIPLIPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60

Qy 61 TWPVKLGWLTPKGEELIAYLGHYQRQRLVADGLLAKKGCPQSGQVAIIADVDERTRKTGE 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 TWPVKLGWLTPKGEELIAYLGHYQRQRLVADGLLAKKGCPQSGQVAIIADVDERTRKTGE 120

Qy 121 AFAAGLAPDCALTVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH 180
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AFAAGLAPDCALTVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH 180

Qy 181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLT 240
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLT 240

Qy 241 EIFLLQQAQQMPEPGWGKITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIMAA 300
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 EIFLLQQAQQMPEPGWGKITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIMAA 300

Qy 301 LDNHFQFKQAYGVTLPDCVLFTAVDCQLNSANSLALFLNWLTPQLSNTPPVSYCFVTSEW 360
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 LDNHFQFKQAYGVTLPDCVLFTAVDCQLNSANSLALFLNWLTPQLSNTPPVSYCFVTSEW 360

Qy 361 RNLSDNSQWTQVSLVFQTTLQQMRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQ 420

RESULT 2
US 09/418,928 2
; Sequence 2, Application US/09418928
; Patent No. 6194740
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/09/418,928
; CURRENT FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 09/291,941
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 08/259,214
; EARLIER FILING DATE: 1994-06-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US 09/418,928 2

Query Match 100.0%; Score 2302; DB 4; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.4e-234;
Matches 440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKAIIPFFSLLIPLIPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60
Db 1 MKAIIPFFSLLIPLIPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60

Qy 61 IWPVKLGWITPKGEELIAYLGHYQRQRLVADGLLAKKGCPQSGQVAIIADVDERTRKTGE 120
Db 61 IWPVKLGWITPKGEELIAYLGHYQRQRLVADGLLAKKGCPQSGQVAIIADVDERTRKTGE 120

Qy 121 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH 180
Db 121 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH 180

Qy 181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSAENVSLTGAVSLASMLT 240
Db 181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSAENVSLTGAVSLASMLT 240

Qy 241 EIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIMAA 300
Db 241 EIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIMAA 300

Qy 301 LTPHPPQKQAYGVTLPTSVLFIAGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFEKW 360
Db 301 LTPHPPQKQAYGVTLPTSVLFIAGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFEKW 360

Qy 361 RRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNTLMTDVKLLYSGTPIFNDVVKTDAKFTLA 420
Db 361 RRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNTLMTDVKLLYSGTPIFNDVVKTDAKFTLA 420

Qy 421 IVNEAKIPAVSLRSHHHHHH 440
Db 421 IVNEAKIPAVSLRSHHHHHH 440

RESULT 3
US 09/291,941 2
; Sequence 2, Application US/09291941A
; Patent No. 6190897
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: Novel PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/09/291,941A
; CURRENT FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 08/259,214
; EARLIER FILING DATE: 1994-06-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US 09/291,941 2

Query Match 100.0%; Score 2302; DB 4; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.4e-234;
Matches 440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKAIIPFFSLLIPLIPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60
Db 1 MKAIIPFFSLLIPLIPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60

Qy 61 IWPVKLGWITPKGEELIAYLGHYQRQRLVADGLLAKKGCPQSGQVAIIADVDERTRKTGE 120
Db 61 IWPVKLGWITPKGEELIAYLGHYQRQRLVADGLLAKKGCPQSGQVAIIADVDERTRKTGE 120

Qy 121 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH 180
Db 121 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH 180

Qy 181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSAENVSLTGAVSLASMLT 240
Db 181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSAENVSLTGAVSLASMLT 240

Qy 241 EIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIMAA 300
Db 241 EIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIMAA 300

Qy 301 LTPHPPQKQAYGVTLPTSVLFIAGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFEKW 360
Db 301 LTPHPPQKQAYGVTLPTSVLFIAGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFEKW 360

Qy 361 RRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNTLMTDVKLLYSGTPIFNDVVKTDAKFTLA 420
Db 361 RRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNTLMTDVKLLYSGTPIFNDVVKTDAKFTLA 420

Qy 421 IVNEAKIPAVSLRSHHHHHH 440
Db 421 IVNEAKIPAVSLRSHHHHHH 440

RESULT 4
US 08/910,798 2
; Application No. US/08910798
; Patent No. 5827697
; GENERAL INFORMATION:
; APPLICANT: KRETZ
; TITLE OF INVENTION: NOVEL PHYTASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:

```
.       MEDIUM TYPE:  Floppy disk
.       COMPUTER:  IBM PC compatible
.       OPERATING SYSTEM:  PC DOS/MS DOS
.       SOFTWARE:  PatentIn Release #1.0, Version #1.25
.     CURRENT APPLICATION DATA:
.       APPLICATION NUMBER:  08/868,435
.       FILING DATE:  August 13,1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BATTE, PHILIP, LISA A.
.       REGISTRATION NUMBER:  38,747
.       REFERENCE/DOCKET NUMBER:  [illegible]
;     TELECOMMUNICATION INFORMATION:
.       TELEPHONE:  619/678 5020
;       TELEFAX:  619/678 5099
.   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
.       LENGTH:  423 amino acids
.       TYPE:  amino acid
.       TOPOLOGY:  linear
.     MOLECULE TYPE:  protein
US-08-868-435-2

  Query Match          96.4%;  Score 2194.5;  DB 4;  Length 423;
  Best Local Similarity  96.1%;  Pred. No. 6e-224;
  Matches  423;  Conservative    0;  Mismatches    0;  Indels    17;  Gaps    1;

Qy       1 MKAIFVLKGWVQPTGSIKFETETDSPSVWDVKKMLVAELQQLKVEHPDMPLFLYDEEGDKVFIGDDFPWEDFVQCVRRLRYLKGS 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKAIFVLKGWVQPTGSIKFETETDSPSVWDVKKMLVAELQQLKVEHPDMPLFLYDEEGDKVFIGDDFPWEDFVQCVRRLRYLKGS 80

Qy      81 IWPVFLSWLFPPVSGLIAYFSHYYGPGWLVAVLLAAVFVLLSAVELAWFLSWRQIN 136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 IWPVFLSWLFPPVSGLIAYFSHYYGPGWLVAVLLAAVFVLLSAVELAWFLSWRQIN 136

Qy     137 AFAATGASFCAFIVSLFVLGFYSNFFDFGVFLSGNANVTAFLFATVSAFFTGLIS 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 AFAATGASFCAFIVSLFVLGFYSNFFDFGVFLSGNANVTAFLFATVSAFFTGLIS 192

Qy     193 EQTAFFSSDFLENFDRTVVLGMNVILAILCLAFVVAINWVLIAAMLVSVAEGSR 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 EQTAFFSSDFLENFDRTVVLGMNVILAILCLAFVVAINWVLIAAMLVSVAEGSR 248

Qy     249 EIFELQQAQGMPEFPWGRFLLDSHQAWNTLGLGLGNAGFYLLGFPFFVAGSAAFLLDLGMAA 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 EIFELQQAQGMPEFPWGRFLLDSHQAWNTLGLGLGNAGFYLLGFPFFVAGSAAFLLDLGMAA 308

Qy     309 EIPHEFQKQAYGVLFQCVLFAHGIKNLANGIGAGIWDCVSDLFVSGRVFLGW 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     309 EIPHEFQKQAYGVLFQCVLFAHGIKNLANGIGAGIWDCVSDLFVSGRVFLGW 360

Qy     361 RESLENGSWSLVGLVFLILGMPSKRFLQLNEFPGPVKRLLAVSLGSGAL 410
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 RESLENGSWSLVGLVFLILGMPSKRFLQLNEFPGPVKRLLAVSLGSGAL 410

Qy     411 IVNDAPLPACSLESCHHHHHHH 433
           |||||||||||||||||||||||
Db     411 IVNDAPLPACSLESCHHHHHHH 433


RESULT  5
US-08-868-435-33
; Sequence 33, Application No/0868435
; Patent No, 6201221
.   GENERAL INFORMATION:
.     APPLICANT:  Van Loon, Antiphius
;     APPLICANT:  Mitchell, David
;     TITLE OF INVENTION:  POLYPEPTIDES WITH PHYTASE ACTIVITY
;     NUMBER OF SEQUENCES:  45
.     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Hoffmann La Roche Inc.
;       STREET:  340 Kingsland Street

;       CITY:  Nutley
;       STATE:  New Jersey
;       COUNTRY:  United States of America
;       ZIP:  07110
;     COMPUTER READABLE FORM:
.       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
.       APPLICATION NUMBER:  08/868,435
.       FILING DATE:
.       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/744,241
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kass, Alan P
;       REGISTRATION NUMBER:  32142
;       REFERENCE/DOCKET NUMBER:  Case Docket 9439
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (201) 235 4205
;       TELEFAX:  (201) 235 2363
;   INFORMATION FOR SEQ ID NO:  34:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  465 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
.     MOLECULE TYPE:  protein
;     FEATURE:
.       NAME/KEY:  misc_feature
;       LOCATION:  104
.       OTHER INFORMATION:  /note "potential N-glycosylation site"
.     FEATURE:
.       NAME/KEY:  misc_feature
;       LOCATION:  119
.       OTHER INFORMATION:  /note "potential N-glycosylation site"
;     FEATURE:
.       NAME/KEY:  misc_feature
;       LOCATION:  205
.       OTHER INFORMATION:  /note "potential N-glycosylation site"
;     FEATURE:
.       NAME/KEY:  misc_feature
;       LOCATION:  228
.       OTHER INFORMATION:  /note "potential N-glycosylation site"
;     FEATURE:
.       NAME/KEY:  misc_feature
;       LOCATION:  337
.       OTHER INFORMATION:  /note "potential N-glycosylation site"
;     FEATURE:
.       NAME/KEY:  misc_feature
;       LOCATION:  474
.       OTHER INFORMATION:  /note "potential N-glycosylation site"
US-08-868-435-34

  Query Match          5.2%;  Score 119;  DB 4;  Length 465;
  Best Local Similarity  21.7%;  Pred. No. 0.0006;
  Matches  102;  Conservative   70;  Mismatches  184;  Indels  114;  Gaps   24;

Qy       7 PELSLEIPETPQSAFAQSEFEEKLESVVIVSREFVKAPI     SAIQIMQLVIELA   78
           ||  ||   |:  |  :   : :: : |:|||| ||      |  :|:  :  :|
Db      52 PFESLEDELSVSSKEFK   I+KEIIVQVLSREGAEYEIASKESKEYSKEIV AIQANAIEQ 111

Qy      79        WPIWPVKLSW  LIPRGGELIAYIG  HYQRQRLVAIKILAKEGPTQGSVAII 130
                    |:   ||   ||| | :  :   |  ||| : :|  ::             |
Db     112 KGKEAFLKIYNYEIGAADELIPFGERQQQLVNSGIESFYQRYKAIARSVVP         FIP 170

Qy     131 ADVDERIRKIGEAFAAGLS-   -   -       APDCAIIVHLQADFSSEDFLENFKKIV 184
           |   :|   :|| | |                   ||  :: :  :|||  :     | |  ||
Db     171 ASSGTHRVIAKSEKFTGFGQAKLAFPGAINKAAFAIGVIL PEGEIEN     NEIHEV 231

Qy     185 QGLDNANVIIAIGSAYQGI AEGIIELEPVLNEFQSNLTRFFEQLIQVII 234
```

RESULT 6
US-08-758-213-1
; Sequence 1, Application US/08758213
; Patent No. 5764498
; GENERAL INFORMATION:
; APPLICANT: LaVallie, E.
; APPLICANT: Jakob, T.
; TITLE OF INVENTION: [illegible]
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Michael A. Mann, P.A.
; STREET: Post Office Box 7908
; CITY: Columbia
; STATE: South Carolina
; COUNTRY: United States
; ZIP: 29202
; COMPUTER READABLE FORM: On attached diskette
; MEDIUM TYPE: Diskette 3.5 inch, 1.4 MB storage
; COMPUTER: Apple Macintosh IM
; OPERATING SYSTEM: Macintosh IM
; SOFTWARE: Microsoft Word 5.1 IM
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,213
; FILING DATE: 27-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/650,659
; FILING DATE: 20-MAY-1996
; APPLICATION NUMBER: 08/489,091
; FILING DATE: 09/Jun/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Mann, Michael A.
; REGISTRATION NUMBER: [illegible]
; REFERENCE/DOCKET NUMBER: [illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (803) 254-8472
; TELEFAX: [illegible]
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 amino acids
; TYPE: Amino acid
; STRANDEDNESS: Not Relevant Applicable
; TOPOLOGY: Linear
US-08-758-213-1

Query Match [illegible]; Score [illegible]; DB 2; Length 386;
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches 87; Conservative 64; Mismatches 129; Indels 112; Gaps 15;

RESULT 7
US-08-692-787-48
; Sequence 48, Application US/08692787
; Patent No. 5882864
; GENERAL INFORMATION:
; APPLICANT: Au, Gane
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: [illegible]
; TITLE OF INVENTION: [illegible]
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: FastSEQ for Windows #1.0, Version #1.00
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/692,787
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Corder, Timothy S.
; REGISTRATION NUMBER: 40,414
; REFERENCE/DOCKET NUMBER: [illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 886 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-692-787-48

```
  Query Match             5.1%;  Score 112;  DB 2;  Length 588;
  Best Local Similarity   29.2%;  Pred. No. 0.00098;
  Matches   87;  Conservative   64;  Mismatches  178;  Indels  112;  Gaps   15;

Qy        8 AIL...                                                    61
            :: : : :: :      :: : :  :  : : :: :                    :
Db      13 SLSLGFELLFPWLDRSVLAK-    -KLKFVILVERHGDRSPI              DIEPID  56

Qy       62          -WEVKLGWLFPPGGELIAYLGHYQKQPEVAIFGLAKKGTPGK    GQVAIIADVD  112
                     || : | | | |  | | | | |:|          :|  :|     || | :
Db       57 PTKESSWPQGFFKQLQLQLGMEQIYELGEYIRKRY-      PKFLNESYKHEQVYIRSTDV  109

Qy      113 ERTRKTCEAFAAGLAPECAITVHTQADTSSPDPLFNPLKTGVCQLDNANVTEAILSRAGG  172
            :|| :      | | |   :::         :|  |: |:        |              :
Db      110 DRTIMSAMTNLAALFPPEGVSIW       NPTLLWQPIPVHTVPLSE                  150

Qy      173 STADFLGHRQTAFRFLEPVNFTFCNDSKEFPKGIFSRLAESCFKVSDSLDAVHGNVYSVS  232
                        :::|  | | | | :: :      ::      |     | :: :
Db      151          -      -- DQILYLPFRN CPRFQELESETLKSEEFQKRLHEYKEFIATLSK  193

Qy      233 VS           LASMLTEIFLLQQAQQMEEFGWGRITDSHQWNTLLSLHNAQFYLLQRTP  282
            :|           : | : :     :      | |       :   |  |     |  :
Db      194 LSGLHGQDLFGIWSKVYDPLYCESVHNFTLPSWATEDTMKKLRELSELSLLSLYGIHKQK  253

Qy      283 EVAKSKATPLLDLIMAALTPHPPQKQAYGVTLPTSVLFIAGHDTNLANLGGALEL  NWTL  342
            | :| :   |:: |: :       |:|  :      ::  : ||| :: :  |:: | |
Db      254 EKTKLQGVEFVNFMYLNGM       KEAIQPICSYKKLIMYSANDKLVPLSGCQVEGQGRVVMKLL  307

Qy      343 PGQPDNTPVSGELVFEKWRELSRNGEQWIQVSLVLQSLQMRGKTPLSLNTPPGSVKLTLA  402
            |  | :  :  | ||:          | : ::                |  |
Db      308 E  EYASCHLIELYFFKKGEYF       VEMYYRN             -EIQKEDPYPLMLP  345

Qy      403 SCPERNAQGMCSLAGFTQIV 422
            ||          | |  | ::|
Db      346 GCSP -   SCPLERFAELV 360


RESULT  9
US-09-097-199-38
; Sequence 38, Appl. #'s 09/097,199
; Patent No. 6218529
;  GENERAL INFORMATION:
;     APPLICANT:  Au, Gino
;     APPLICANT:  Black, D. Mark
;     APPLICANT:  Balph, David
;     APPLICANT:  Vollett, Robert
;     TITLE OF INVENTION:  BPH MARKERS AND TARGETS FOR DIAGNOSIS,
;     TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;     NUMBER OF SEQUENCES:  87
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC Compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Patentin Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  09/097,199
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/xxx,xxx
;       FILING DATE:  31-JUL-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nikolchew, Richard A.
;       REGISTRATION NUMBER:  P-42,523
;       REFERENCE/DOCKET NUMBER:  DBC-0018
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (512) 474-7577
;     INFORMATION FOR SEQ ID NO:  38:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  386 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-097-199-38


  Query Match             5.1%;  Score 112;  DB 4;  Length 386;
  Best Local Similarity   29.2%;  Pred. No. 0.00098;
  Matches   87;  Conservative   64;  Mismatches  178;  Indels  112;  Gaps   15;

Qy        8 AIL...                                                    61
            :: : : :: :      :: : :  :  : : :: :                    :
Db      13 SLSLGFELLFPWLDRSVLAK-    -KLKFVILVERHGDRSPI              DIEPID  56

Qy       62          -WEVKLGWLFPPGGELIAYLGHYQKQPEVAIFGLAKKGTPGK    GQVAIIADVD  112
                     || : | | | |  | | | | |:|          :|  :|     || | :
Db       57 PTKESSWPQGFFKQLQLQLGMEQIYELGEYIRKRY-      PKFLNESYKHEQVYIRSTDV  109

Qy      113 ERTRKTCEAFAAGLAPECAITVHTQADTSSPDPLFNPLKTGVCQLDNANVTEAILSRAGG  172
            :|| :      | | |   :::         :|  |: |:        |              :
Db      110 DRTIMSAMTNLAALFPPEGVSIW       NPTLLWQPIPVHTVPLSE                  150

Qy      173 STADFLGHRQTAFRFLEPVNFTFCNDSKEFPKGIFSRLAESCFKVSDSLDAVHGNVYSVS  232
                        :::|  | | | | :: :      ::      |     | :: :
Db      151          -      -- DQILYLPFRN CPRFQELESETLKSEEFQKRLHEYKEFIATLSK  193

Qy      233 VS           LASMLTEIFLLQQAQQMEEFGWGRITDSHQWNTLLSLHNAQFYLLQRTP  282
            :|           : | : :     :      | |       :   |  |     |  :
Db      194 LSGLHGQDLFGIWSKVYDPLYCESVHNFTLPSWATEDTMKKLRELSELSLLSLYGIHKQK  253

Qy      283 EVAKSKATPLLDLIMAALTPHPPQKQAYGVTLPTSVLFIAGHDTNLANLGGALEL  NWTL  342
            | :| :   |:: |: :       |:|  :      ::  : ||| :: :  |:: | |
Db      254 EKTKLQGVEFVNFMYLNGM       KEAIQPICSYKKLIMYSANDKLVPLSGCQVEGQGRVVMKLL  307

Qy      343 PGQPDNTPVSGELVFEKWRELSRNGEQWIQVSLVLQSLQMRGKTPLSLNTPPGSVKLTLA  402
            |  | :  :  | ||:          | : ::                |  |
Db      308 E  EYASCHLIELYFFKKGEYF       VEMYYRN             -EIQKEDPYPLMLP  345

Qy      403 SCPERNAQGMCSLAGFTQIV 422
            ||          | |  | ::|
Db      346 GCSP -   SCPLERFAELV 360


RESULT  7
US-09-146-283-2
; Sequence 2, Application US/09146283
; Patent No. 5976546
;  GENERAL INFORMATION:
;     APPLICANT:  Lass, Reinet
;     APPLICANT:  Maag, Curtis L.
;     APPLICANT:  Wu, Hongyu
;     TITLE OF INVENTION:  Immunostimulatory Compositions
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Dehlinger & Associates
;       STREET:  350 Cambridge Ave, Suite 250
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94306
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Patentin Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
```

RESULT 2
US/08/579,823A/2
; Sequence 2, Application US/08/579,823A
; Patent No. 6080409
; GENERAL INFORMATION:
; APPLICANT: Laos, Keiter
; APPLICANT: Kuo, Chris L.
; APPLICANT: Wu, Bonnie
; TITLE OF INVENTION: Immunostimulatory Oligonucleotides and Method
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 700 Hansen Way
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/579,823A
; FILING DATE: 03-DEC-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7646-0010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE: Prostatic acid phosphatase-GM-CSF
; INDIVIDUAL ISOLATE: fusion protein; Fig. 1
US 09-146-283-2

Query Match 5.1%; Score 117; DB 2; Length 515;
Best Local Similarity 19.6%; Pred. No. 0.0018;
Matches 87; Conservative 64; Mismatches 178; Indels 112; Gaps 15;

Qy 8 AIIIFFLVIIIFLIIGSAVQGFEEVELDEVVGVGQLWGDLRVQMLGAVPLDGWFT 61
Db 13 MGSPLPSLLLFSMLWLDLSVLAK--GLKEVIHVEPRHGDVSPI--VH--GTIFPLD 59

Qy 62 WRVKGMGLIFKDDELGAYLSHKGMRPSAEIVAMMEDLAKRKTMLG--SLVATSVL 116
Db 57 PIFPPSWPLFVILIGMELPHIHDPNAPRGLKEFLRKNSVPEDVGTLKLIV 109

Qy 117 ERTKKTGHAFAASLAFLAIIVHFQAFRSMDNPELRDVGLNGVYDAIMSAVS 172
Db 110 DRTLMSAMTNLAALFPPEGVSIW------NPILLWQPIPVHTVPLSE--- 150

Qy 173 GIAGGQTASLLLSVLGAFGSNLGFKKEAKQMLSGTLQALESFLRVGAQNVSLIGA 232
Db 151 ----------DQLLYLPPRN CPRFQELESETLESFFEEYREHIEGFGNLQV 195

Qy 233 VSLASMLTTIITLLGGAGMPTGWGKTTTGHGWNTLLSGLRNAGFTLLGWTP 283
Db 196 LSGLHGQDLEGIWSKVTEEIVTLSEDGVRGFTEEAWALTQSIMIKELSELIALRN 255

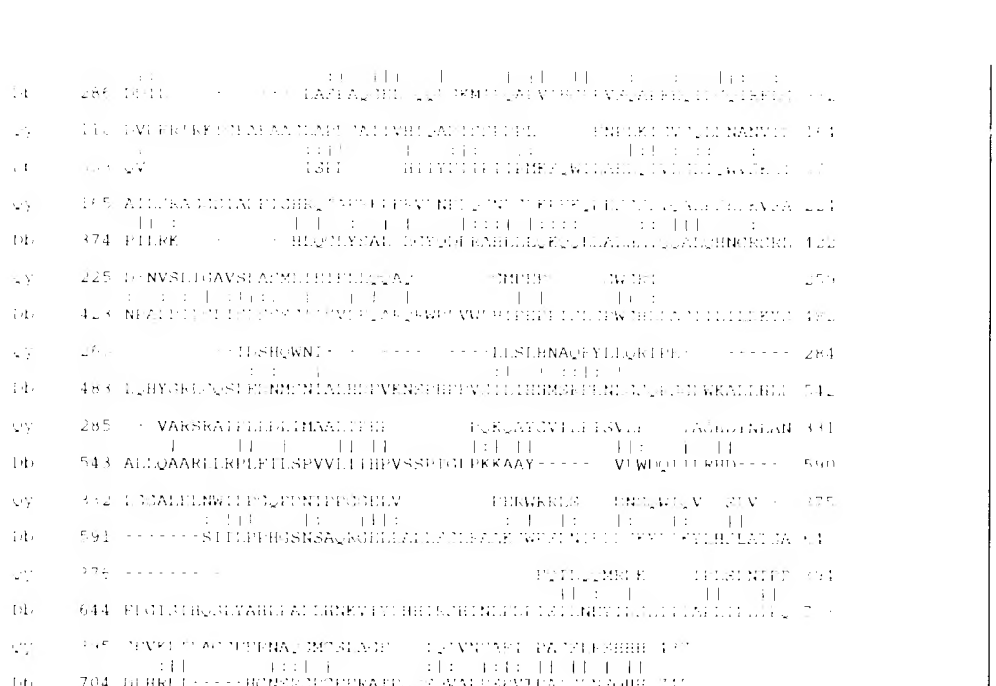
Qy 284 PVARSNALPLTLAGMAALTHPLGRGAVGVTLDSVLFIAGHDTNLANLGGALEL NWTL 342
Db 256 EEAPLGIVLVNELMDN EGALGLSYGHQTISVSARSVFWPMMEGGKDDVPLLA 307

Qy 343 PNGPNPFIHELVFLHWFDLGVNGSWPCVSLVGLGLQPRSEPFTLTLNPHGVTLLA 402
Db 308 FTACTHLDHTYFEKDL VLEVTPAN - - EGLHFYFELMEP 345

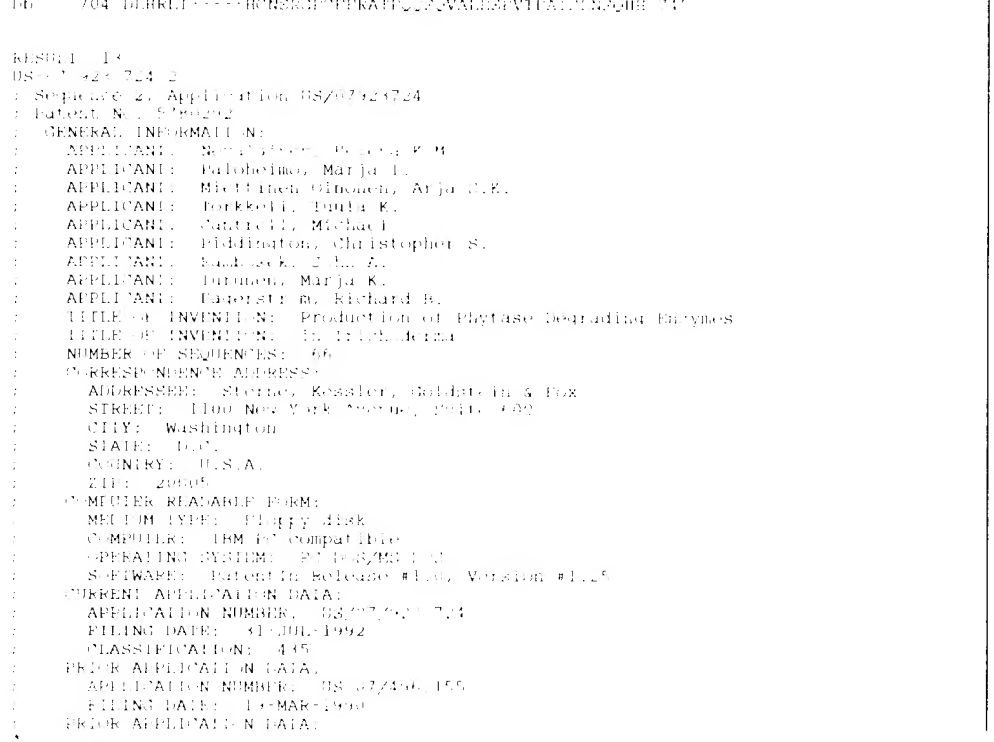
Qy 403 GPHERNAGMGSLAGPLGIV 422
Db 346 GCSF GCTLERPAELV 360

RESULT 11

```
US-09-580-515-2
; Sequence 2, Application US/09/580515
; Patent No. 6210662
;  GENERAL INFORMATION:
;       APPLICANT: Lans, Reiner
;                  Rusyn, Curtis L.
;                  Wu, Baoliu
;       TITLE OF INVENTION: Immunostimulatory Compositions
;       NUMBER OF SEQUENCES: 10
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Testa, Jr. & Associates
;            STREET: 450 Cambridge Ave, Suite 250
;            CITY: Palo Alto
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 94306
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: Patentin Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/580,515
;            FILING DATE: 24-May-2000
;            CLASSIFICATION: Unknown
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/xxx,xxx
;            FILING DATE: 03-SEP-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Testa, Hitchcock
;            REGISTRATION NUMBER: 42,702
;            REFERENCE/DOCKET NUMBER: 2696.0010.21
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650 324 0880
;            TELEFAX: 650 324 0960
;  INFORMATION FOR SEQ ID NO: 2:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 515 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Homo sapiens
;            INDIVIDUAL ISOLATE: Prostatic acid phosphatase-GM-CSF
;                                fusion protein; Fig. 1
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-344-195-2


  Query Match             5.1%;  Score 117;  DB 1;  Length 515;
  Best Local Similarity   19.8%;  Pred. No. 0.0013;
  Matches   87;  Conservative   63;  Mismatches  178;  Indels  112;  Gaps   15;

Qy        3 AIIIPPESLIIPLIPqQSAIAQSEPELKIESVVIVSRHGVRAPTKATQLMQDVTPDAWPT- 61
            ::: : :: ::    :  ::     :: : : :: :: :  :           : :::  :
Db       14 SLSQGFELLIEFDYTFTSKEHGQLIAK      ELKEVGIEYETHGSYSRS-------SGTPFR- 56

Qy       62         WPVKTGWLIIPLQQEELIAYLGHYQRQRLVAEHKLAKGSYQqAIVATIADVD 112
                    ||   : ||  : :     || : ::          ::   ::    || : :
Db       57 PTKESSWPQnSPQQLIQLFQMSQLGIYLSFCKEYERKFY       PPSLNLSYKSSQVYFRSTDV 109

Qy      113 EETRKTSAIAAGLAIGLFIPVIIQAVICSFDPRFTNTIKSVQQNSANVTIATISSAQGG 172
            :::  :  : : :  :::          :: : :: :       :
Db      110 DRTLMSAMTNSLAALFPPGVTIW       NPTIIHEVETYLDH                153

Qy      173 SIAPTTNLYGLIAFSLAEVLNFQNSNICLKKIKNNSVQTVDLSYSLNLVKSAQNAVPLDGA 232
                    :::  : : : :: :      ::      |    : :: :
Db      154                 DqvLYLPTFRN CPRFQELESELLKSEFEQKRIPQYSPWEIATIGH 199

Qy      233 VS          LACMLGLFILLLQQAIQMPEPQWVRFIQQLVNCLPSLIPKETVQETT 282
            :|          : : :  :      : :|      :  : :  :  : : :
Db      200 LSGLHGQDDLFGIWSKVYDPLYCESVHNFTLPSWATEDTMTKLRELSELSLLSVYGIHKQP 259

Qy      283 LVAGSVAFTLSSLMAALFIHFLQKAVGVTLTGVLTVVGNLIVLNVFVSLWLKIK 337
            | :|  :   |:: |:  :       |:|  :      :  : :|| :: |  |:| : :
Db      260 EVMSSAATPQDLFDYTWSFYRDLYKRVSGVLQGPALGAPLVASLILVFLMSLAGTGR 315

Qy      338 PQQPDNIPPGSELVFIFDWRRSLSDNSQWIQVSLVFQFIAQGMKKFIFLSLNIPPSFVFIIA 397
            | | :   || ||:              | : ::              |      | :
Db      316 PYASCHLIELYFFKGEYF  --  --VEMYYRN              FIQHIEYFLMSI  417

Qy      398 SSFLFNAGMGSLAGFLQIV 422
            ||      | |  | :::
Db      418 GCSP      SCFLERFAELV 460


RESULT  12
PCT/US95/04910-14
; Sequence 14, Application PC/TUS9504910
;  GENERAL INFORMATION:
;    APPLICANT:  The Government of the United
;    APPLICANT:  States of America as represented
;    APPLICANT:  by the Secretary, Department of
;    APPLICANT:  Health and Human Services
;    TITLE OF INVENTION:  CHARACTERIZATION OF A NEWLY DISCOVERED
;    TITLE OF INVENTION:  LYMPHOTROPIC VIRUS AND THE USE OF SUCH VIRU
;    TITLE OF INVENTION:  COMPONENTS THEREOF IN DIAGNOSTIC ASSAYS,
;    TITLE OF INVENTION:  AND VACCINES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04910
;      FILING DATE:  21-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US08/231,526
;      FILING DATE:  22-APR-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILLIAM S. FEILER
;      REGISTRATION NUMBER:  26,728
;      REFERENCE/DOCKET NUMBER:  2926.4125PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  995 amino acids
;      TYPE:  amino acids
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
PCT/US95/04910-14


  Query Match             4.8%;  Score 106;  DB 5;  Length ??;
  Best Local Similarity   19.9%;  Pred. No. 0.057;
  Matches  117;  Conservative   66;  Mismatches  184;  Indels  ...1;  Gaps   24;

Qy       7 PELSLIIP      - --- --      SLIPQSAEAQSEPELKLESVVIVSRHGVRAPTKA 57
           |:  ||                    : ||  |  |    : :   |:|      ||     :
Db     227 EYTATTIL_PQDVRQIGALEAYAGTVLT_  QPENGTTLPLQ_LAAVFGQDELPSRGAEQ 281

Qy      58 QLVSELGQAWQKWRVKLGWLPRRGAEAIYGHVQANLDKLVAEFGLLASNLTGGVMIDA 115
```



RESULT 13
US-07-923-724-2
; Sequence 2, Application US/07923724
; Patent No. 5290702
; GENERAL INFORMATION:
; APPLICANT: [illegible]
; APPLICANT: Paloheimo, Marja I.
; APPLICANT: Miettinen-Oinonen, Arja S.K.
; APPLICANT: Korkkoff, Tuula K.
; APPLICANT: Cantrell, Michael
; APPLICANT: Piddington, Christopher S.
; APPLICANT: Rambosek, John A.
; APPLICANT: Jarmuen, Marja K.
; APPLICANT: Paloheimo M. Richard B.
; TITLE OF INVENTION: Production of Phytase Degrading Enzymes
; TITLE OF INVENTION: in Filamentous Fungi
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W. Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/923,724
; FILING DATE: 31-JUL-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/856,155
; FILING DATE: 19-MAR-1992
; PRIOR APPLICATION DATA:



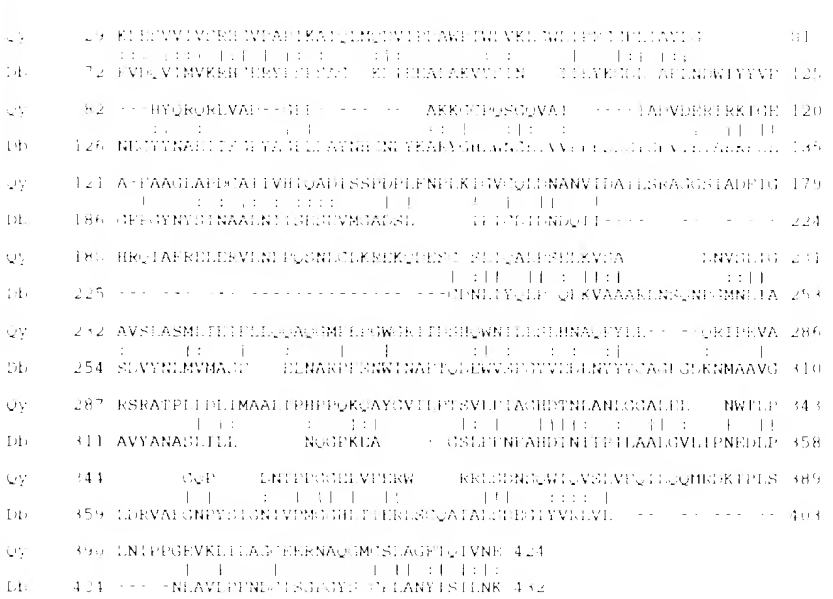
; APPLICATION NUMBER: US 07/844,077
; FILING DATE: 27-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8610600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1758.0240004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-923-724-2

Query Match 4.4%; Score 101.5; DB 1; Length 479;
Best Local Similarity 19.6%; Pred. No. 0.05;
Matches 96; Conservative 56; Mismatches 157; Indels 153; Gaps 23;

RESULT 14
US-08-609-426A-2
; Sequence 2, Application US/08609426A
; Patent No. 5830743
; GENERAL INFORMATION:
; APPLICANT: Rosalaion, Rocelia K.M.
; APPLICANT: Paloheimo, Marja I.
; APPLICANT: Miettinen-Oinonen, Arja S.K.
; APPLICANT: Korkkoff, Tuula K.
; APPLICANT: Cantrell, Michael
; APPLICANT: Piddington, Christopher S.
; APPLICANT: Rambosek, John A.
; APPLICANT: Jarmuen, Marja K.
; APPLICANT: Paloheimo M. Richard B.
; APPLICANT: Boraston, Christine S.

```
;  TITLE OF INVENTION:  Production of Phytase Controlled Eukaryotes
;  TITLE OF INVENTION:  in Trichoderma
;  NUMBER OF SEQUENCES:  69
;  CORRESPONDENCE ADDRESS:
;    ADDRESSEE:  Stevens, Kessler, Goldstein & Fox
;    STREET:  1155 New York Avenue, Suite 500
;    CITY:  Washington
;    STATE:  D.C.
;    COUNTRY:  U.S.A.
;    ZIP:  20005
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE:  Floppy disk
;    COMPUTER:  IBM PC compatible
;    OPERATING SYSTEM:  PC-DOS/MS-DOS
;    SOFTWARE:  PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER:  US/08/474,626A
;    FILING DATE:  01-MAR-1995
;    CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  US 07/925,404
;    FILING DATE:  31-JUL-1992
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  US 07/766,155
;    FILING DATE:  19-MAR-1991
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  US 07/711,422
;    FILING DATE:  29-APR-1987
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  DK 9601/86
;    FILING DATE:  29-APR-1986
;  ATTORNEY/AGENT INFORMATION:
;    NAME:  Reed, Grant E.
;    REGISTRATION NUMBER:  41,294
;    REFERENCE/DOCKET NUMBER:  1050.0070001
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE:  (202) 371-2600
;    TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;  SEQUENCE CHARACTERISTICS:
;    LENGTH:  479 amino acids
;    TYPE:  amino acid
;    TOPOLOGY:  linear
;  MOLECULE TYPE:  protein
US-08-609-426A-2


  Query Match              4.1%;  Score 191.5;  DB 2;  Length 479;
  Best Local Similarity   19.6%;  Pred. No. 0.0%;
  Matches   89;  Conservative   56;  Mismatches  154;  Indels  153;  Gaps   23;

Qy       29 KLESVVIVSPHCVFALKAEIDMPLYKLLAKEIWDRLDGKPNELVATLA        81
            :: |:|:|:|:|:|:  :|:     :  :     |  |:|:||
Db       73 EVDQVFHHVPFPCDPYTPCDAS  EDCEDALARVTGTD   TEEPKSS APLDWFYYVF 125

Qy       82     QYQDQMLVAD--GLL           AEKGTPQVQVAF       IAIVEFEPLPKFCE 129
                :|  :       |||           |:| .|  :           | | :| :|
Db      126 NETYYNAFTLDYYAKSIAPNRCNGTRAFYGHGTQLSVVFTPSYSYSTEAVKRFCE 181

Qy      130 ATAAGLAIGTATVRDQHDLSFLMNRLRDYPEQLGNGEKVMQGLNANVIHALERPAGSTAPEG 189
            |  |  | | |:|:  :|  |  |  |  |  |
Db      182 QFPSYNYCINAALNEIGLSFVMPGAFSE     IPLPDDDNDQII                224

Qy      190 HRQIAFKPLERVENPQQNEIKPPEKQDESF QEIQAFLSPLRVGA        DNVSLIG 231
                                        |:|| |: : |:|:         :|||
Db      225                              CDNLTVQLP QFKVAAAFLNNQNPQMNLTA 254

Qy      232 AVSLASHLIEFELLQADAMFFTSWCRETLCHLWNLLLTLDNAQFYLL     QRTPEVA 286
            :  :  :  :  :  |  |  |  :|:  :  :  :|  :  :  |
Db      255 SDVYNLMVNMAQT   ELNAPLPCNWCNATLQELMVSEYVEPLNVTFYAIL DGKNMAAVG 310

Qy      287 RSPATLTKTMMAEIRRLKEQKAFNWTLDWLRLNMYGIANVTNRER   NWTLP 333
                |  | |:      :  |:|       |:|  |:||:  :|  :  |  |
```

```
Db      311 AVYGNLMVNMAQTELNAPLPCNWCNATLQELMVSEYVEPLNVTFYAILDGKNMAAVG 370

Qy      334 AVSLLSHLIEFELLQADAMFFTSWCRETLCHLWNLLLTLDNAQFYLLQRTPEVA 388
            :  :  :  :  :  |  |  |  :|:  :  :  :|  :  :  |
Db      371 SDVYNLMVNMAQTELNAPLPCNWCNATLQELMVSEYVEPLNVTFYAILDGKNMAAVG 430

Qy      389 LNEPPSEVKLEIAVETEERNAQDMESLACETQIVNE 424
            |  |  |  :  |  |  |  |:|:|:|:|:|:|:
Db      431 LDNVSLIGNIAVSHLIEFELLQADAMFFTSWCRETLCHLWNLLLTLDNAQFYLLQRTPEVA 490

RESULT  15
US-08-474-626-4
; Sequence 4, Application US/08/474,626
; Patent No. 5861286
;  GENERAL INFORMATION:
;    APPLICANT:  NEVALAINEN, HEIKKA K.M.
;    APPLICANT:  PAPPINEMO, MARJA T.
;    APPLICANT:  PAPERSTROM, RICHARD B.
;    APPLICANT:  MEITTINEN-NINEN, ARJA S.
;    APPLICANT:  PUNNEN, MARJA K.
;    APPLICANT:  KANERSEK, JOHN A.
;    APPLICANT:  PETERSTON, CHRISTOPHER S.
;    APPLICANT:  BOWSTON, CHRISTINE S.
;    APPLICANT:  CANTRELL, MICHAEL A.
;    TITLE OF INVENTION:  RECOMBINANT CELLS, DNA CONSTRUCTS
;    TITLE OF INVENTION:  VECTORS AND METHODS FOR EXPRESSING ENZYMES IN GALENI
;    TITLE OF INVENTION:  ENZYMES IN DESIRED RATIOS
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STEVENS, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1155 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,626
;      FILING DATE:  24-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/925,404
;      FILING DATE:  27-JUL-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/925,404
;      FILING DATE:  31-JUL-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  REED, GRANT E.
;      REGISTRATION NUMBER:  41,294
;      REFERENCE/DOCKET NUMBER:  1050.0070001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202 371 2600
;      TELEFAX:  202 371 2540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  479 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
US-08-474-626-4


  Query Match              4.4%;  Score 191.5;  DB 2;  Length 479;
  Best Local Similarity   19.6%;  Pred. No. 0.0%;
  Matches   89;  Conservative   56;  Mismatches  154;  Indels  153;  Gaps   23;
```



Search completed: November 20, 2001, 10:00:45
Job time: 122 sec

SeqSearch version 1.0
Copyright (c) 1993 - 2000 Compugen Ltd.

Comparison matrix: BLOSUM62
Gap penalty: 10 / 0.05
Processor load during scan: 0%

Run on: November 23, 2001, 09:51:06 ; Search time 111.7 Seconds
(without alignments)
1161.780 Million cell updates/sec

Title: US-09-580-515-2
Perfect score: 2302
Sequence: 1 MKAIIPPTSIIPIIPSA.....IVNEARIFAVSRSPHHHHHH 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112676 seqs, 50062008 residues

Total number of hits satisfying chosen parameters: 112676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A:Genseqp00dt:*
1: /cgn2_6/ptodata/genseqp/genseqp00dt_*
2: /cgn2_6/ptodata/genseqp/genseqp00dt_*
3: /cgn2_6/ptodata/genseqp/genseqp00dt_*
4: /cgn2_6/ptodata/genseqp/genseqp00dt_*
5: /cgn2_6/ptodata/genseqp/genseqp00dt_*
6: /cgn2_6/ptodata/genseqp/genseqp00dt_*
7: /cgn2_6/ptodata/genseqp/genseqp00dt_*
8: /cgn2_6/ptodata/genseqp/genseqp00dt_*
9: /cgn2_6/ptodata/genseqp/genseqp00dt_*
10: /cgn2_6/ptodata/genseqp/genseqp00dt_*
11: /cgn2_6/ptodata/genseqp/genseqp00dt_*
12: /cgn2_6/ptodata/genseqp/genseqp00dt_*
13: /cgn2_6/ptodata/genseqp/genseqp00dt_*
14: /cgn2_6/ptodata/genseqp/genseqp00dt_*
15: /cgn2_6/ptodata/genseqp/genseqp00dt_*
16: /cgn2_6/ptodata/genseqp/genseqp00dt_*
17: /cgn2_6/ptodata/genseqp/genseqp00dt_*
18: /cgn2_6/ptodata/genseqp/genseqp00dt_*
19: /cgn2_6/ptodata/genseqp/genseqp00dt_*
20: /cgn2_6/ptodata/genseqp/genseqp00dt_*
21: /cgn2_6/ptodata/genseqp/genseqp00dt_*
22: /cgn2_6/ptodata/genseqp/genseqp00dt_*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2302	100.0	449	22	AAB37892	Escherichia coli ph
2	2295	99.7	442	21	AAB36267	Ismir2/APPA plasmid
3	2295	99.7	442	21	AAB36261	E15/APPA plasmid t
4	2295	99.7	442	21	AAB36259	E15/APPA plasmid t
5	2295	99.7	442	21	AAB36260	E15/APPA plasmid t
6	2295	99.7	442	21	AAB36261	E15/APPA plasmid t
7	2295	99.7	442	21	AAB36262	SV40/APPA plasmid
8	2295	99.7	442	21	AAB36263	Ismir2/APPA plasmid
9	2221	96.5	442	21	AAY94753	E. coli acidic pho
10	2194.5	95.3	424	20	AAY01513	An Escherichia col
11	119	5.2	446	21	AAY69549	Aspergillus fumiga
12	119	5.2	465	19	AAW81456	Aspergillus fumiga
13	119	5.2	465	20	AAY89905	A. fumigatus phyta
14	118	5.1	440	21	AAB20507	Aspergillus fumiga
15	118	5.1	440	21	AAY69550	Aspergillus fumiga
16	117	5.1	386	19	AAW57418	Protein sequence f
17	117	5.1	386	20	AAW95081	Protein sequence f
18	117	5.1	386	21	AAY59293	Fragment of acid ph
19	117	5.1	515	18	AAW19762	EAP-CM 251 immunoge
20	111	4.8	440	21	AAB20508	Aspergillus fumiga
21	111	4.8	440	21	AAY69551	Aspergillus fumiga
22	111	4.8	467	20	AAY43171	A. fumigatus phyta
23	111	4.8	467	21	AAB20528	Aspergillus fumiga
24	111	4.8	467	21	AAY69570	Mutant Aspergillus
25	110	4.8	440	21	AAB20509	Aspergillus fumiga
26	110	4.8	440	21	AAY69552	Aspergillus fumiga
27	110	4.8	467	21	AAY69574	Mutant Aspergillus
28	106.5	4.6	1048	20	AAW97720	Staphylococcus aur
29	106	4.6	995	16	AAR84658	SIIVpat: p-pol prot
30	105	4.6	22	21	AAB36255	Bacterial AEPA pro
31	102.5	4.5	385	20	AAW38543	Mouse protein seque
32	101.5	4.4	479	15	AAR46791	pH 2.5 acid phosph
33	101.5	4.4	479	15	AAR46235	A. niger pH 2.5 ac
34	100	4.3	1374	19	AAW69753	Herpes simplex vir
35	99.5	4.3	466	17	AAR88623	Chimeric phytase.
36	98.5	4.3	5542	21	AAB23752	S. cerevisiae his4
37	97	4.2	297	21	AAY48064	Arabidopsis thalia
38	97	4.2	422	21	AAB20517	Paxillus involutus
39	97	4.2	442	19	AAW62869	Paxillus involutus
40	97	4.2	442	20	AAY89903	P. involutus phyta
41	97	4.2	1410	21	AAG31536	Arabidopsis thalia
42	97	4.2	1415	21	AAG31538	Arabidopsis thalia
43	97	4.2	1445	21	AAG31537	Arabidopsis thalia
44	96.5	4.2	427	21	AAB09969	Human lysophosphat
45	96	4.2	1.9e+03	21	AAY13868	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAB37892
ID AAB37892 standard; protein; 449 AA.
XX
AC AAB37892;
XX
DT 07-MAR-2001 (first entry)
XX
DE Escherichia coli phytase enzyme.
XX
KW Escherichia coli; phytase enzyme; modelling; phytate digestion;
KW nutrition.
XX
OS Escherichia coli.
XX
PN WO200071728-A1.
XX
PD 30-NOV-2000.
XX
PF 25-MAY-2000; 2000WO-US14846.
XX
PR 25-MAY-1999; 99US-0318528.
XX
PA (GIVE) DIVERSA CORP.
XX
PI Short PI; F; 449; FV;
XX
DR WPI; 2001-112993/12.
DR N-PSDB; AAC88885.
XX
PT Improving the nutritional value of phytate containing feedstuffs - by
PT phytase enzymes which catalyze the liberation of inorganic phosphate
PT from the phytates
XX

```
cS   Claim 1; Fig 1; 11pp; English.
XX
CC   The present sequence is a phytase enzyme from Escherichia coli B.
CC   the enzyme catalyses the liberation of inorganic phosphate from the
CC   phytate in phytate containing foodstuffs and can thus be used to improve
CC   the nutritional value of phytate rich ingredients.
XX
SQ   Sequence  440 AA;


  Query Match             Score 2902;  DB 2;  Length 440;
  Best Local Similarity   Pred. No. 3.2e-224;
  Matches  440;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MRAILIVLLSLFTVAAAQSAPAFDKIKSEPELMSDAYNVLWFTQSAMSNLVFSLQGIGWDS  60
         | | | ||||| |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKQTLIPLFSLLLPLTPPSTTAQSEPQLKLESVEIVSRHGVRAPTKATQLMQDVTPDAWP  60

Qy      61 IWQVKLGWLTPKGMQLIALLGGFYRQKFQSQGLLAKGSCPTPNSIYVWADVDQRTLKTGE  120
           | ||| |||||||| | |||||||||||||| |||||||||||||||||||||||||||||||||||||
Db      61 TWPAPLQVKTKPDNELSPYDLVIVPTPMEKKLKDSVFLQIAETRALPLSQTAAGVQIQAYW  120

Qy     121 AEAAAGAPVVALVDVLTHMVKNTTNSLCQRTDALCPVTRLDTDVKESAFCQAGPKFMPSAL  180
           | | |||| | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATAAAQAPDEAITVTDQSLNSEPDSLFDQAKKYFQDHRNRVSKEFADIFTHSEIVNLGYGL  180

Qy     181 SLEDKVKPFFELLNQTINFDLTGKPVPDRYHVYDMGENVNFMLWLFNCHNTVQPPMSGGTA  240
           | | |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TLTDQVSSALVSLMQLAKFDTTTKLYFLSAHDTTIAGLQMALDVYNGLLPPYASNLTFELY  240

Qy     241 EIVFKNGDKDVDVSYIFCQNETQLRGLTPVTLETYVSGITKRSEFCSQCTLETFCSLAKIV  300
           | | |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 EIVLDKQGQEPFVRMLLNDRVLQLPGCPEDTLCPLEKFQQIVDGSIEPRCWHEEVAQGDVC  300

Qy     301 LTPHPPQKAVSVTLPSYVHIAEENLANLSGALEFNIIEGKAKQFVDLYKSALPFGSGHEKA  360
           | | |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 LTPHPPQKAVSVTLPSYVHIAEENLANLSGALEFNIIEGKAKQFVDLYKSALPFGSGHEKA  360

Qy     361 KGGSGNSLVTPAVSLVGLVFLSVSFLLGSL  391
           | | |||| |||||||||||||||||||||||
Db     361 PGSGNNSQWTQVSGLVTPTEPPMPLTLSLM  391

Qy     121 IVNEARIPASFLASFPDITKAA 143
           | | |||| |||| | | | |||
Db     421 IVNEARIPASFLGSDGNETAA 443


RESULT   2
AAB36257
ID   AAB36257 standard; protein; 442 AA.
XX
AC   AAB36257;
XX
DT   25-FEB-2001  (first entry)
XX
DE   DAMA2/AcPA plasmid translated sequence.
XX
KW   transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW   environmental pollution; pig.
XX
OS   Mus musculus - chimeric.
OS   Escherichia coli - chimeric.
XX
RN   [1]
RP   1-442
RX   PATENT_NO-WO200048477-A1.
XX
RA   Forsberg C.W., Golovan S., Phillips J.P.;
XX
RT   Transgenic nonhuman animal for gastrointestinal tract specific
RT   expression of a protein, preferably phytase, comprises a nucleic acid
RT   sequence including a heterologous transgene construct encoding the
RT   protein.
XX
RL   Disclosure; Fig 5; 152pp; English.
XX
CC   The present invention provides transgenic animals which produce secreted
CC   proteins, in this case pigs which expresses phytase in the salivary
CC   gland, how phytase production levels result in phytate in the feed being
CC   excreted and causing phosphorus contamination in water, as well as
CC   reducing the growth of animals. The invention provides a nucleic acid
CC   transgenes containing the E. coli APPA phytase coding sequence.
XX
SQ   Sequence  442 AA;


  Query Match             Score 2902;  DB 2;  Length 442;
  Best Local Similarity   99.5%;  Pred. No. 3.2e-224;
  Matches  440;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 MRAILIVLLSLFTVAAAQSAPAFDKIKSEPELMSDAYNVLWFTQSAMSNLVFSLQGIGWDS  60
           | | |||| |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKQTLIPLFSLLLPLTPPSTTAQSEPQLKLESVEIVSRHGVRAPTKATQLMQDVTPDAWP  60

Qy      61 IWQVKLGWLTPKGMQLIALLGGFYRQKFQSQGLLAKGSCPTPNSIYVWADVDQRTLKTGE  120
           | ||| |||||||| | |||||||||||||| |||||||||||||||||||||||||||||||||||||
Db      61 TWPAPLQVKTKPDNELSPYDLVIVPTPMEKKLKDSVFLQIAETRALPLSQTAAGVQIQAYW  120

Qy     121 AEAAAGAPVVALVDVLTHMVKNTTNSLCQRTDALCPVTRLDTDVKESAFCQAGPKFMPSAL  180
           | | |||| | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATAAAQAPDEAITVTDQSLNSEPDSLFDQAKKYFQDHRNRVSKEFADIFTHSEIVNLGYGL  180

Qy     181 SLEDKVKPFFELLNQTINFDLTGKPVPDRYHVYDMGENVNFMLWLFNCHNTVQPPMSGGTA  240
           | | |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TLTDQVSSALVSLMQLAKFDTTTKLYFLSAHDTTIAGLQMALDVYNGLLPPYASNLTFELY  240

Qy     241 EIVFKNGDKDVDVSYIFCQNETQLRGLTPVTLETYVSGITKRSEFCSQCTLETFCSLAKIV  300
           | | |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 EIVLDKQGQEPFVRMLLNDRVLQLPGCPEDTLCPLEKFQQIVDGSIEPRCWHEEVAQGDVC  300

Qy     301 LTPHPPQKAVSVTLPSYVHIAEENLANLSGALEFNIIEGKAKQFVDLYKSALPFGSGHEKA  360
           | | |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 LTPHPPQKAVSVTLPSYVHIAEENLANLSGALEFNIIEGKAKQFVDLYKSALPFGSGHEKA  360

Qy     361 KGGSGNSLVTPAVSLVGLVFLSVSFLLGSL  391
           | | |||| |||||||||||||||||||||||
Db     361 PGSGNNSQWTQVSGLVTPTEPPMPLTLSLM  391

Qy     121 IVNEARIPASFLASFPDITKAA 143
           | | |||| |||| | | | |||
Db     421 IVNEARIPASFLGSDGNETAA 443


RESULT   3
AAB36258
ID   AAB36258 standard; protein; 442 AA.
XX
AC   AAB36258;
XX
DT   20-FEB-2001  (first entry)
XX
DE   K1%/AcPA plasmid translated sequence.
XX
KW   transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW   environmental pollution; pig.
XX
OS   Rattus sp. - chimeric.
```

ID [illegible] chimeric.
XX
AC [illegible]
XX
DT [illegible]
XX
DE [illegible]
XX
KW [illegible]
XX
OS [illegible]
XX
PN [illegible]
XX
PD [illegible]
XX
PF [illegible]
XX
PR [illegible]
XX
PA [illegible]
XX
PT [illegible]
XX
DR [illegible]
XX
PT Transgenic non human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein
XX
PS Disclosure; Fig 19; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence.
XX
SQ Sequence [illegible] AA;

Query Match [illegible]; Score 2235; DB 21; Length [illegible];
Best Local Similarity [illegible]; Pred. No. 2.1e-217;
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

RESULT 1
AAB66259
ID AAB66259 standard; Protein; 432 AA.
XX
AC AAB66259;
XX
DT 20-FEB-2001 (first entry)
XX
DE Rl5/APPA plasmid translated sequence.
XX
KW Transgenic animal; salivary protein; phytase; phosphorus excretion; [illegible];
KW environmental pollution; pig.
XX
OS Rattus sp. chimeric.
OS Escherichia coli chimeric.
XX
PN WO-[illegible]-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA00430.
XX
PR 23-APR-1999; 99US-0130508.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI; 2000-687245/67.
DR N-PSDB; AAC68296.
XX
PT Transgenic non human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein
XX
PS Disclosure; Fig 19; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence.
XX
SQ Sequence 432 AA;

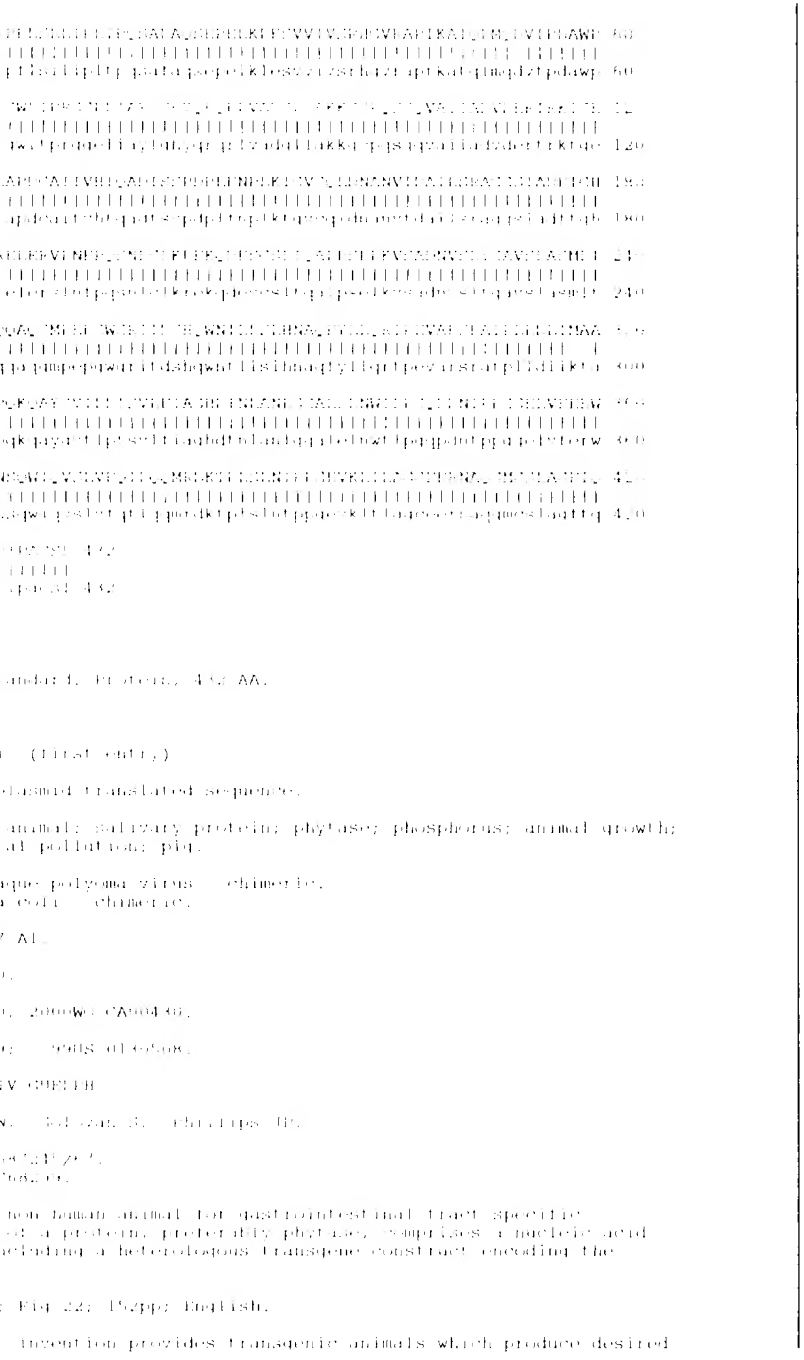
Query Match [illegible]; Score 2235; DB 21; Length [illegible];
Best Local Similarity [illegible]; Pred. No. 2.1e-217;
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

RESULT 5
AAB68260
ID AAB68260 standard; Protein; 432 AA.
XX
AC AAB68260;
XX
DT 20-FEB-2001 (first entry)
XX
DE 6F5/AFPA plasmid translated sequence.
XX
KW transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
OS Rattus sp. - chimeric.
OS Escherichia coli - chimeric.
XX
PN WO200061247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000512.
XX
PR 23-APR-1999; 99US-0130508.
XX
PA (UYGU) UNIV GUELPH.
XX
PI Forsberg CW; Golovan S; Phillips JP;
XX
DR WPI; 2000-687245/67.
DR N-PSDB; AAB68297.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein
XX
PS Disclosure; Fig 20; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli AFPA phytase coding sequence.
XX
SQ Sequence 432 AA;

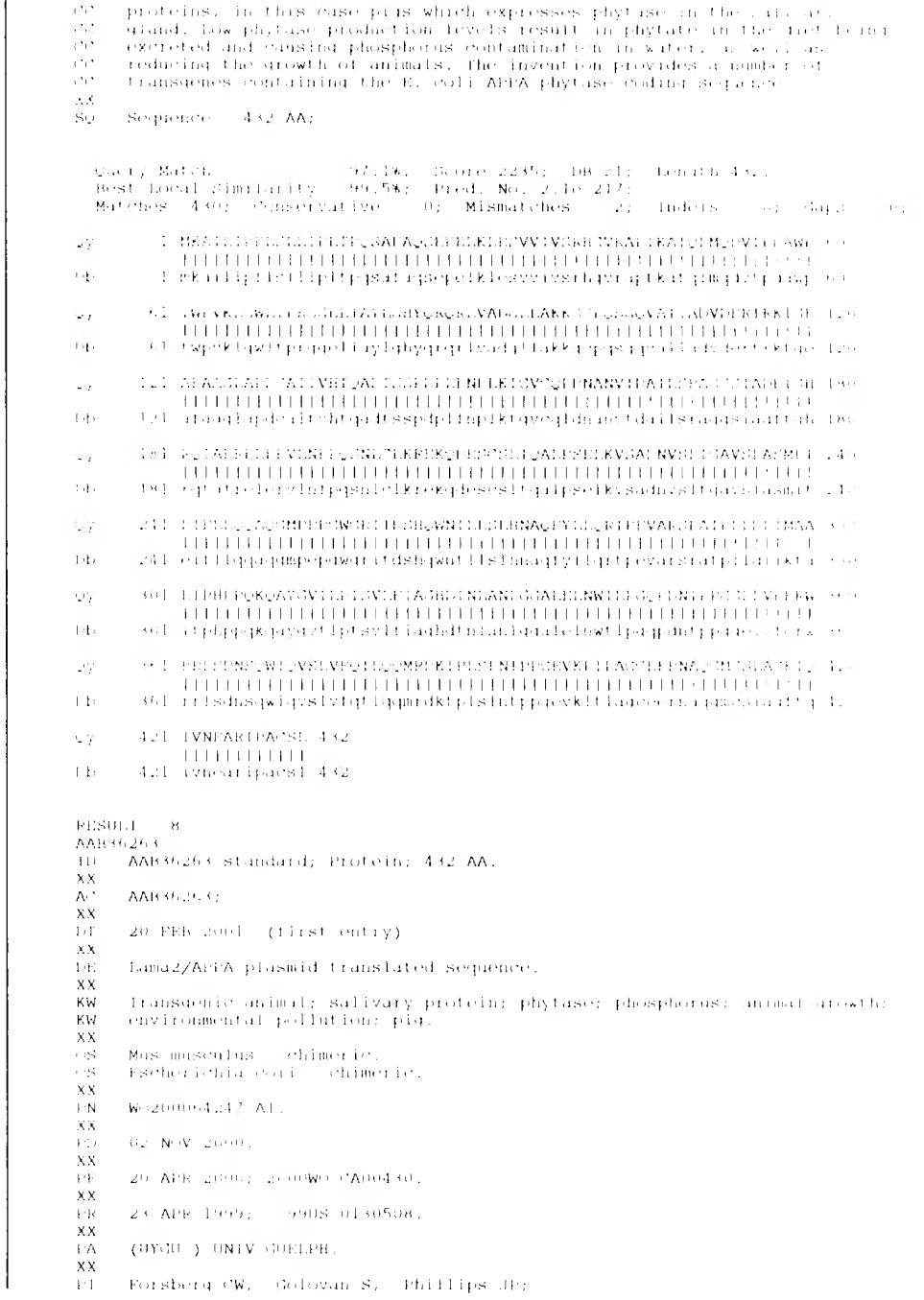
Query Match 97.1%; Score 2245; DB 21; Length 432;
Best Local Similarity 99.5%; Prod. No. 2.1e-217;
Matches 430; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 6
AAB68261
ID AAB68261 standard; Protein; 432 AA.
XX
AC AAB68261;
XX
DT 20-FEB-2001 (first entry)
XX
DE 6F7/AFPA plasmid translated sequence.
XX
KW transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
OS Rattus sp. - chimeric.
OS Escherichia coli - chimeric.
XX
PN WO200061247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000512.
XX
PR 23-APR-1999; 99US-0130508.
XX
PA (UYGU) UNIV GUELPH.
XX
PI Forsberg CW; Golovan S; Phillips JP;
XX
DR WPI; 2000-687245/67.
DR N-PSDB; AAB68298.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein
XX
PS Disclosure; Fig 21; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli AFPA phytase coding sequence.
XX
SQ Sequence 432 AA;

Query Match 97.1%; Score 2245; DB 21; Length 432;
Best Local Similarity 99.5%; Prod. No. 2.1e-217;
Matches 430; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



RESULT 7
AAB36263
ID AAB36263 standard; Protein; 432 AA.
XX
AC AAB36263;
XX
DT 20-FEB-2001 (first entry)
XX
DE SV40/APPA plasmid translated sequence.
XX
KW transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
OS Rhesus macaque polyoma virus - chimeric.
OS Escherichia coli - chimeric.
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA00430.
XX
PR 23-APR-1999; 99US-0130508.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;



RESULT 8
AAB36263
ID AAB36263 standard; Protein; 432 AA.
XX
AC AAB36263;
XX
DT 20-FEB-2001 (first entry)
XX
DE Lama2/APPA plasmid translated sequence.
XX
KW transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
OS Mus musculus - chimeric.
OS Escherichia coli - chimeric.
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA00430.
XX
PR 23-APR-1999; 99US-0130508.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;

```
XX
PN   WO2001079745/2
DR   N_PSDB; AAY68500.
XX
PT   Transgenic non-human animal for gastrointestinal tract specific
PT   expression of a phytase, preferably phytase, comprises a nucleic acid
PT   sequence including a heterologous transgene construct containing the
PT   protein
XX
PS   Disclosure; 134 pp; English.
XX
CC   The present invention provides transgenic animals which produce desired
CC   proteins, in this case preferably phytase in the salivary
CC   gland. Low phytase production levels result in phytate in the diet being
CC   excreted and causing phosphorus contamination in water, as well as
CC   reducing the growth of animals. The invention provides a number of
CC   transgenes containing the E. coli AppA phytase coding sequence.
XX
SQ   Sequence  432 AA;


  Query Match             97.1%;  Score 2221;  DB 21;  Length 432;
  Best Local Similarity   99.5%;  Pred. No. 2.1e-214;
  Matches  430;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 MKAIIIFILSLLIIPLVAQMAQTEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MKAIIIFILSLLIIPLVAQMAQTEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP  60

Qy       61 TWPVKLGWLTPKGEESMIDSGIRFRKYFNSLDFLPPEYNPSEVWIRSTDVDRTLKSGEA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 TWPVKLGWLTPKGEESMIDSGIRFRKYFNSLDFLPPEYNPSEVWIRSTDVDRTLKSGEA  120

Qy      121 AFAAGLAPDCALTVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 AFAAGLAPDCALTVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH  180

Qy      181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLT  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLT  240

Qy      241 EIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIKTA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 EIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIKTA  300

Qy      301 LIPHPQQKQAYGVTLPTSVLFIAGHDTNLANLSGALELNWTLPGQPDNTPPGGELVFERW  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 LIPHPQQKQAYGVTLPTSVLFIAGHDTNLANLSGALELNWTLPGQPDNTPPGGELVFERW  360

Qy      361 RRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNTLMCPVHDVKLTASPFLRNAQEMSLAGFLQ  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 RRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNTLMCPVHDVKLTASPFLRNAQEMSLAGFLQ  420

Qy      421 IVNEAKIPACSL 432
            ||||||||||||
Db      421 IVNEAKIPACSL 432


RESULT   9
AAY64753
ID   AAY64753 standard; Protein; 432 AA.
XX
AC   AAY64753;
XX
DT   12-FEB-2001  (first entry)
XX
DE   E. coli acidic phosphatase appA2 fragment amino acid sequence.
XX
KW   Phosphatase; phytase activity; pH 6.0; acidic phosphatase; appA2;
KW   periplasmic phosphoanhydride phosphohydrolase; animal feed.
XX
OS   Escherichia coli.
XX
```

```
OS   Escherichia coli.
XX
PD   05-JUL-2000.
XX
PF   21-MAR-2000;  2000WO-US007500.
XX
PR   21-MAR-1999;  99US-0125332.
XX
PA   (PHAR ) PHARMACIA & UPJOHN INC.
XX
PI   Lei X;
XX
DR   WPI; 2001-279759/29.
DR   N_PSDB; AAY64753.
XX
PT   Novel phosphatase having improved phytase activity useful as an additive
PT   feed for improving the accessibility of phosphate to an animal -
PT   produced by treating phosphatase with protease -
XX
PS   Claim 27; Fig 6; 48pp; English.
XX
CC   This invention relates to a phosphatase fragment with improved phytase
CC   activity. The phosphatase fragment is created by treating the phosphatase
CC   with a protease. The invention includes an Escherichia coli nucleotide
CC   sequence which encodes an acidic phosphatase appA2 which has improved
CC   phosphatase activity. The appA gene of Escherichia coli was originally
CC   defined as E. coli periplasmic phosphoanhydride phosphohydrolase,
CC   although it was not previously known to have phytase activity. The
CC   phosphatase fragment can be used in animal feed, to improve the
CC   accessibility of phosphate to the animal. The phosphatase gene which
CC   encodes the protein can be used in the production of phytase. The present
CC   sequence represents the appA2 protein of the invention.
XX
SQ   Sequence  432 AA;


  Query Match             96.5%;  Score 2221;  DB 21;  Length 432;
  Best Local Similarity   99.5%;  Pred. No. 2.5e-216;
  Matches  430;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 MKAIIIFILSLLIIPLVAQMAQTEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MKAIIIFILSLLIIPLVAQMAQTEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP  60

Qy       61 TWPVKLGWLTPKGEESMIDSGIRFRKYFNSLDFLPPEYNPSEVWIRSTDVDRTLKSGEA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 TWPVKLGWLTPKGEESMIDSGIRFRKYFNSLDFLPPEYNPSEVWIRSTDVDRTLKSGEA  120

Qy      121 AFAAGLAPDCALTVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 AFAAGLAPDCALTVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGH  180

Qy      181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLT  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 RQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLT  240

Qy      241 EIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIKTA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 EIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIKTA  300

Qy      301 LIPHPQQKQAYGVTLPTSVLFIAGHDTNLANLSGALELNWTLPGQPDNTPPGGELVFERW  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 LIPHPQQKQAYGVTLPTSVLFIAGHDTNLANLSGALELNWTLPGQPDNTPPGGELVFERW  360

Qy      361 RRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNTLMCPVHDVKLTASPFLRNAQEMSLAGFLQ  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 RRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNTLMCPVHDVKLTASPFLRNAQEMSLAGFLQ  420

Qy      421 IVNEAKIPACSL 432
            ||||||||||||
Db      421 IVNEAKIPACSL 432
```

```
RESULT  10
AAY02513
ID   AAY02513 standard; protein; 423 AA
XX
AC   AAY02513;
XX
DT   27 MAY 1999  (first entry)
XX
DE   An Escherichia coli phytase protein.
XX
KW   Escherichia coli B; phytase; animal feed composition;
KW   phytate degradation.
XX
OS   Escherichia coli.
XX
PN   WO9916878-A1.
XX
PD   8-APR-1999.
XX
PF   24-SEP-1998;  98WO-US019736.
XX
PR   26-SEP-1997;  97US-0059240.
XX
PA   (GIVE ) GIVEN IMAGING LTD.
XX
PI   Kressel K;
XX
DR   WPI; 1999-286641/24.
DR   N-PSDB; AAX29540.
XX
PT   New substantially pure phytase polypeptide derived from Escherichia
PT   coli B   useful in an animal feed composition for degrading phytate
XX
PS   Claim 1; Fig 1; 45pp; English.
XX
CC   The present sequence represents an Escherichia coli B phytase
CC   polypeptides. The phytase forms an animal feed composition for
CC   degrading phytate, where the enzyme hydrolyses the
CC   phospho-mono-ester bond in phytate.
XX
SQ   Sequence  423 AA;


  Query Match             95.8%;  Score 2164.5;  DB 2;  Length 423;
  Best Local Similarity   96.1%;  Pred. No. 2.9e-214;
  Matches  423;  Conservative    0;  Mismatches    0;  Indels    17;  Gaps    1;

Qy        1 MKAIIIFTLSILPLIFTQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MKAIIIFTLSILPLIFTQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60

Qy       61 IWEVKNEWEPEYGQIIAYISGMYQRQMYLAQKLLAKGFLPFQGSVAFIADVDERTRKTGE 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 IWEVKNEWEPEYGQIIAYISGMYQRQMYLAQKLLAKGFLPFQGSVAFIADVDERTRKTGE 120

Qy      121 ALAAGLAGLFAPINDSDTAEAKSSHTPDLEQVDEQYQSVVDKYAGQTMDRLNQLTGGLSL 180
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121                     QADTSSPDPLFNPLKTGVCQLDNMNVTDAILSRAGGSIADFTGH 163

Qy      181 DGLWSLVTMEKNVGEPWFNETTVSSVWQVDNLFDSKFGLTKQGRKNRFVMAGNQLQGASFE 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      164 QQTAIRELEEVIGKPEVPESERGPVILAKVLSNDSTYLEFSISLAQEGGAEKMFIPAFGGM 223

Qy      241 FTLDGVWSMISFMSVQEIIPLSTEAPKEAPQWNEFDHSEKWNTLISLHNAQFYLWMAEAAP 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      224 YTLDPVQPAAPQPQTPKSTFDHSQVNTLISLHNAQFYLWMAEAAP 268

Qy      301 LTGPLKTHLKDSVSLVFMDSVSQLHEFYSSTMRNQYPSKPGGKKGTDFILPQWFTADYTMR 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      269 LTGPLKTHLKDSVSLVFMDSVSQLHEFYSSTMRNQYPSKPGGKKGTDFILPQWFTADYTMR 328

Qy      361 DFTDHSKWLTYGAKISMFSSGDMDKLFGMAGRLNPEDKWRLENKTSEKHYGSTVPFIANNF 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      329 DFTDHSKWLTYGAKISMFSSGDMDKLFGMAGRLNPEDKWRLENKTSEKHYGSTVPFIANNF 388

Qy      421 IVNLAKIEAVSLKSHHHHHH 440
            ||||||||||||||||||||
Db      389 FVNLAKIEAVSLKSHHHHHH 408


RESULT  11
AAY69549
ID   AAY69549 standard; protein; 440 AA.
XX
AC   AAY69549;
XX
DT   19-APR-2000  (first entry)
XX
DE   Aspergillus fumigatus ATCC 13073 mature phytase.
XX
KW   Phytase; myo-inositol hexakisphosphate phosphohydrolase; starch liquefaction;
KW   thermostable; animal feed; monogastric animal; phytate phosphorus;
KW   phosphate availability.
XX
OS   Aspergillus fumigatus ATCC13073.
XX
PN   EP969089-A1.
XX
PD   05-JAN-2000.
XX
PF   23-JUN-1999;  99EP-0111949.
XX
PR   29-JUN-1998;  98EP-0111960.
XX
PA   (HOFF ) HOFFMANN-LA ROCHE & CO AG F.
XX
PI   Brugger R,  Lehmann M,  Wyss M;
XX
DR   WPI; 2000-094129/09.
XX
PT   New stabilized enzyme formulation, useful for feed compositions for
PT   monogastric animals
XX
PS   Example 3; Fig 14; 101pp; English.
XX
CC   The invention relates to a novel stabilised dry or liquid enzyme
CC   formulation, comprising phytase (myo-inositol hexakisphosphate
CC   phosphohydrolase) and one or more stabilising agents including
CC   xylitol or ribitol; polyethylene glycols with a molecular weight of 200
CC   to 4000 Da, preferably 1000 to 3350 Da; the disodium salts of malonic,
CC   glutaric and succinic acid; carboxymethylcellulose; and sodium alginate.
CC   The stabilised phytase formulation is used in a method for preparing a
CC   feed composition for monogastric animals (e.g., pigs, poultry) and
CC   provides a monogastric animal with its dietary requirements of
CC   phosphorus. Although a large amount of phosphate is present in animal
CC   feed in the form of phytate phosphorus, monogastric animals are unable
CC   to utilise this form of phosphate, resulting in the addition of extra
CC   phosphate to the feed of such animals. Phytase enhances the nutritional
CC   value of plant material without the need for adding additional phosphate
CC   to the feed. The level of phosphate pollution in the environment is
CC   reduced by adding phytase to animal feed, as the animal can make use of
CC   the inorganic phosphate liberated from phytate phosphorus using the
CC   enzyme. The phytase formulation of the invention has an improved
CC   thermostability and can therefore remain stable during heat-treatment and
CC   non-heat-treatment processing methods such as extrusion, expansion
CC   and pelletising. Sequences AAY69544-Y69556 represent the amino acid
CC   sequences of mature phytases from a variety of fungi (mainly different
CC   strains of Aspergillus species), which were used in an exemplified selection of
CC   the present invention to determine a consensus phytase sequence
CC   designated phytase 1 (AAY69556). An initial consensus sequence (AAY69557)
CC   was determined using the program PRETTY, and the phytase 1 protein was
CC   derived from the initial consensus. Consensus phytase sequences were
CC   produced in the exemplifications, as a prior art consensus phytase
CC   sequence was found to have higher thermal stability than natural
CC   phytases.
XX
```

```
Sq   Sequence   449 AA;

  Query Match          5.2%;  Score 119;  DB 21;  Length 449;
  Best Local Similarity   21.7%;  Pred. No. 1e+03;
  Matches  102;  Conservative   70;  Mismatches  183;  Indels  114;  Gaps   24;

Qy       7 PFLSLLIPLIPQSAFAQSEPELKLESVVIVSRHGVRAPT     KATQLMQDVTPDA--  58
           |  |:   |:  |   :  : ::  | :::||| | ||     |  ::|  :  :|
Db      27 ptlsiqdelsvs-kplk   lgvsl pqlrlayqpt--kgsdtyadwlsaqsnd-  77

Qy       59        WCIWFVFLSW  :::E:IELIAYS-  HS:I:EIVAS PSAEEIP:_:LVAL  108
                   |:  ||  ||  | :  :  |  ||: : :|  ::           |
Db       84 kapkrqlpkyqvtlrlisgie ppvsglkqlgvfkqqsyksladtvpq        131

Qy      109 AIVPFIFFTPEAPAAIL              AIP^AIIVEILAPLSSPGTLFNELEQV  154
            :  :|   :::| | :                ||  ::  :  :::|  :     | |  ||
Db      132 asasdrvlasgekllevtppakladpatnraapaisvli psseqlsv-  -ntldhqv  187

Qy      155 PLPTWATVTAIETAAKIR AIIGHEPLPAVPEIIFVFEIANGLTEFARSPGGLKE  210
            |    |    |   : :  : :|||               | |  ::  : ||     :||
Db      188 r-----  -tkktvasqladevaantt----- -- --altap dlfrarnokhlpqvtl  226

Qy      211 PLPTGKLFVVMRHGDRYPTVSKGKKMEIEGELAFVHQWQAEGALSP  257
            |     | | |: :  |  |  :::|  :  |      :::| | ||
Db      227 dedrvvlkgwlsvdsalkqngvfqgqfrqlpltpqglpqvaalyq  273

Qy      258 PKEKLFIWTSDVDQRTLKTGAAFLQGLFPIEGLNIWNPILLLPLFEAVPVEAETGSRNPGFPVTAFQEAFVLSEHSHFLEKYGDGLTSNITGTALYDIAHTEDKLLRMKNVSFFQLPSLVSENSAFLDVGAGSLYTFTSTETFAAIQNLGLSEDDVDLFMSYSKPKETRIMASSLQPCCPLFQLLKNTEYAQYNTLNVKGTYLEPTQSAGSPLVKDLFHILDQAIQGKSGRLPSYTPEIYILGYSHDTTIAGLQMALDVYNGLLPDYASCHLVELYQNSSNEYFVEMFYNGQEQKMPGCSPVCPLDRFLELTKPVIPQDWSTECMLDNQEEVSITIDGDYLPMQTCTVIQDRILLRYADDLNPGKEDLSSTWV  449

RESULT  12
AAW84459
ID   AAW84459 standard; Protein; 465 AA.
XX
AC   AAW84459;
XX
DT   28-APR-1999  (first entry)
XX
DE   Aspergillus fumigatus phytase enzyme.
XX
KW   Phytase; modified phytase; food additive; phytate; animal manure.
XX
OS   Aspergillus fumigatus.
XX
PN   AU9869943-A.
XX
PD   01-OCT-1998.
XX
PF   25-MAR-1998;   98AU-069943.
XX
PR   25-MAR-1997;   97DE-0012179.
XX
PA   (BOEH ) BOEHRINGER LA MANNH & CO AG.
XX
PI   Kostrewa D;  Kusamoto K;  Tomschy A;  Van Loon A;
PI   Vogel K;  Wyss M;
XX
DR   WPI; 1998-596456/50.
DR   N-PSDB; AAX03334.
XX
PT   New modified phytase with increased activity - produced by
PT   introducing specific amino acid changes based on analysis of
PT   three-dimensional enzyme structures, used as feed and food additive
XX
PS   Example 1; Fig 6; 116pp; English.
XX
CC   The present sequence represents a phytase enzyme. The specification
CC   describes the production of modified phytase with improved properties.
CC   This comprises modelling the three-dimensional structure of the phytase
CC   being modified, and optionally that of another phytase with better
CC   activity, on the basis of the 3-structure of A. niger phytase,
CC   identifying amino acid residues in the active site that differ
CC   [illegible] phytase encoding DNA by
CC   changing at least one of the differing amino acids in the active
CC   [illegible] expressing the DNA, wherein the phytase
CC   are useful as feed additives for converting phytate to myo-inositol
CC   and inorganic phosphate. Addition of modified phytase to food
CC   reduces the amount of phytate in animal manure.
XX
SQ   Sequence   465 AA;

  Query Match          5.2%;  Score 119;  DB 19;  Length 465;
  Best Local Similarity   21.7%;  Pred. No. 0.0034;
  Matches  102;  Conservative   70;  Mismatches  183;  Indels  114;  Gaps   24;

Qy       7 PFLSLLIPLIPQSAFAQSEPELKLESVVIVSRHGVRAPT     KATQLMQDVTPDA--  58
           |  |:   |:  |   :  : ::  | :::||| | ||     |  ::|  :  :|
Db      43 ptlsiqdelsvs-kplk   lgvsl pqlrlayqpt--kgsdtyadwlsaqsnd-  93

Qy      59        WCIWFVFLSW  :::E:IELIAYS-  HS:I:EIVAS PSAEEIP:_:LVAL  108
                  |:  ||  ||  | :  :  |  ||: : :|  ::           |
Db      94 kapkrqlpkyqvtlrlisgie ppvsglkqlgvfkqqsyksladtvpq        141

Qy     109 AIVPFIFFTPEAPAAIL              AIP^AIIVEILAPLSSPGTLFNELEQV  154
           :  :|   :::| | :                ||  ::  :  :::|  :     | |  ||
Db     142 asasdrvlasgekllevtppakladpatnraapaisvli psseqlsv-  -ntldhqv  198

Qy     155 PLPTWATVTAIETAAKIR AIIGHEPLPAVPEIIFVFEIANGLTEFARSPGGLKE  210
           |    |    |   : :  : :|||               | |  ::  : ||     :||
Db     199 r-----  -tkktvasqladevaantt----- -- --altap dlfrarnokhlpqvtl  247

Qy     211 PLPTGKLFVVMRHGDRYPTVSKGKKMEIEGELAFVHQWQAEGALSP  257
           |     | | |: :  |  |  :::|  :  |      :::| | ||
Db     248 dedrvvlkgwlsvdsalkqngvfqgqfrqlpltpqglpqvaalyq  294

Qy     258 PKEKLFIWTSDVDQRTLKTGAAFLQGLFPIEGLNIWNPILLLPLFEAVPVEAETGSRNPGFPVTAFQEAFVLSEHSHFLEKYGDGLTSNITGTALYDIAHTEDKLLRMKNVSFFQLPSLVSENSAFLDVGAGSLYTFTSTETFAAIQNLGLSEDDVDLFMSYSKPKETRIMASSLQPCCPLFQLLKNTEYAQYNTLNVKGTYLEPTQSAGSPLVKDLFHILDQAIQGKSGRLPSYTPEIYILGYSHDTTIAGLQMALDVYNGLLPDYASCHLVELYQNSSNEYFVEMFYNGQEQKMPGCSPVCPLDRFLELTKPVIPQDWSTECMLDNQEEVSITIDGDYLPMQTCTVIQDRILLRYADDLNPGKEDLSSTWV  465

RESULT  13
AAY99905
ID   AAY99905 standard; protein; 465 AA.
XX
AC   AAY99905;
XX
DT   07-DEC-1999  (first entry)
XX
DE   A. fumigatus phytase protein sequence.
XX
KW   Phytase; variant; enzyme; phosphorus liberation; phytase substrate;
KW   phytate level reduction; animal manure; food preparation;
KW   soy processing; inositol manufacture.
XX
OS   Aspergillus fumigatus.
XX
PN   WO9946362-A1.
```

```
XX
DE   6 JUL 2000
XX
DE   21 MAR 2000;   2000WO-DK000154.
XX
DR   23 MAR 1999;   1999DK-0000407.
DR   15 JUN 1999;   1999DK-0000906.
DR   10 SEP 1999;   1999DK-0001276.
DR   21 JAN 2000;   2000DK-0000071.
XX
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Svendsen A;
XX
DR   WPI; 2000-504144/46.
DR   N-PSDB; AAZ27122.
XX
PT   New variant phytase enzymes, used for liberating phosphorus from a
PT   phytase substrate, for reducing phytate levels in animal manure and in
PT   feed and food preparations
XX
PS   Claim 28; Fig 8; 141pp; English.
XX
CC   This sequence represents the Aspergillus fumigatus phytase sequence.
CC   The invention relates to variant phytase enzymes with specific amino acid
CC   substitutions for improved properties. The phytase variants can be used
CC   for liberating phosphorus from a phytase substrate. They can be used for
CC   reducing phytate levels in animal manure. They can be used in feed or
CC   food preparations. The phytase DNA can also be used to produce transgenic
CC   plants which can be used in feeds or foods. The phytase variants can also
CC   be used in soy processing and in the manufacture of inositol or
CC   derivatives. The phytase variants can have altered activities such as pH
CC   stability, temperature stability, pH profile, temperature profile,
CC   specific activity (in particularly in relation to pH and temperature),
CC   substrate specificity, substrate cleavage pattern, substrate binding,
CC   position specificity, the velocity and level of release of phosphate
CC   from corn, reaction rate, phytase degradation rate and end level of
CC   released phosphate reached.
XX
SQ   Sequence   467 AA;
```

```
  Query Match          5.2%;  Score 119;  DB 20;  Length 467;
  Best Local Similarity   21.7%;  Pred. No. 0.0034;
  Matches  102;  Conservative   70;  Mismatches  184;  Indels  114;  Gaps   24;

Qy       7 PELSLLIPLIIPQSAFAQSEPELKLESVVIVSRHGVRAPT     KATQLMQDVTPDA -- 58
           || ||   |:  |   :    : ::  | ::|||| | ||     | ::  :  :|
Db      52 ptlsledelsvvsvklpk-  derlilvqvlsrhgvrapt-sskskkykklatqlpnatdl 108

Qy      59 :       WPIWEVKSGW  LTPKGGELIAYLG  HYQRQRLVADGLLAKKGCPQPGQVAII 108
             |.  ||   ||| | :  :   |  ||| : :|  ::           |
Db     109 kqktatlktynytlcaddtptdeqplvnsnktygpykdhrsvpa     -   tlr 158

Qy     109 ADVDERTRKTGEAFAAGL             - APDCAITVHTQADTSSPDPLFNPLKTGV 154
           |  ::  :|| : |                  :|  ::  : :::|  :     | |  ||
Db     159 aspsdrvtasaqkflegfypqkladpatnvaaptsvsvii psetlp     atldhqv 212

Qy     155 YQLDNANVTDAILSRAGGSI ALPTSWPEVFDRQLAFRFLDIVMNFTQWAQKQDKSRLI 214
           |     |   | : :  : :|||             | |  :  : ||     :||
Db     213 v       vkttvsqplpdcvrantt             altap  dtvtvtvdkhipqvtlt 261

Qy     215 QALDSEL  KVCARNVSLDSAVSVADLFEIFSRDGWPHSAWGSIFCPWNNTDLPL 271
             |  | | |:  |   |  |  ::  :  |  |  |  |  |  |  |  |||
Db     262 dedvvsidwdwvstdtvarisdasqlqpfqplttnav       w     kkyvyfpsl 307

Qy     272 HNAQVEFLQKIEFVARSKATPLLSLIMAALEFHPEQKQ        AYGVELPTSVLFIA 323
           |:|  :  :  :  :  :  :  :  :|| | |  |  |        |  | | :
Db     308 qkyyavvtnspdapapipfttpdcdiarifvspqdtstvtfkvnpatfpldatmyy 365

Qy     324 GHDTNLANLGGALELNWTLPGQPDNTPPGGELVFEKWRRLSD-NSQWI  QVSLVFQ 377
           |  :  :  |        |  : | |       |  : |   :: |:        |:
Db     366 dtsldnsmsvsfttvl       atyntdplsttsvqvskvddpvsnswvmptarrypv 408
```

```
Qy     378 TLQAMNSALETGNIPFGLVKLIAPPVQNAEKYPMFGLVNQAL 1.
           |:|   :| ||  :     :  : | ||   |  :  :|
Db     409 tmpksckqql vralindtvspihpdd dkrrqktlntfkqlswar 1??


RESULT  14
AAB20507
ID   AAB20507 standard; Protein; 410 AA.
XX
AC   AAB20507;
XX
DT   05-DEC-2000  (first entry)
XX
DE   Aspergillus fumigatus 82722 phytase SEQ ID NO:7.
XX
KW   Phytase; acidic; thermostability; mutation; mutagenesis; pH profile;
KW   temperature stability; pH profile; temperature profile; reaction rate;
KW   specific activity; substrate specificity; substrate cleavage pattern;
KW   substrate binding; position specificity; phytate degradation rate;
KW   food; feed; phytate; manure.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200045603-A1.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-DK000025.
XX
PR   22-JAN-1999;   1999DK-0000092.
PR   21-SEP-1999;   1999DK-0001340.
XX
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Lehmann M;
XX
DR   WPI; 2000-504144/46.
XX
PT   Novel phytases with improved properties such as temperature stability,
PT   pH stability and substrate specificity, for use in pharmaceuticals and
PT   compound foods and feeds
XX
PS   Example 1; Fig 1a-d; 240pp; English.
XX
CC   The present invention describes improved phytases, preferably with
CC   increased thermostability, and methods for producing them. The methods
CC   can be used for producing phytases with improved properties e.g.
CC   temperature stability, pH stability, pH profile, temperature profile,
CC   specific activity, substrate specificity, substrate cleavage pattern,
CC   substrate binding, position specificity, the velocity and level of
CC   release of phosphate from corn, reaction rate, phytate degradation rate
CC   and end level of released phosphate. The phytases can be used in pharma
CC   pharmaceutical compositions or compound food or feeds. The feed can be
CC   used to reduce levels of phytate in animal manure, by converting it
CC   into lower inositol phosphates and/or inositol and inorganic phosphate.
CC   The present sequence represents a phytase sequence from the present
CC   invention.
XX
SQ   Sequence   410 AA;
```

```
  Query Match          5.1%;  Score 116;  DB 21;  Length 410;
  Best Local Similarity   21.7%;  Pred. No. 0.0049;
  Matches  102;  Conservative   70;  Mismatches  184;  Indels  114;  Gaps   24;

Qy       7 PELSLLIPLIIPQSAFAQSEPELKLESVVIVSRHGVRAPT     KATQLMQDVTPDA -- 58
           || ||   |:  |   :    : ::  | ::|||| | ||     | ::  :  :|
Db      27 ptlsledelsvvsvklpk-  derlilvqvlsrhgvrapt-sskskkykklatqlpnatdl 83

Qy      59 :       WPIWEVKSGW  LTPKGGELIAYLG  HYQRQRLVADGLLAKKGCPQPGQVAII 108
             |.  ||   ||| | :  :   |  ||| : :|  ::           |
Db      84 kqktatlktynytlcaddtptdeqplvnsnktygpykdhrsvpa     -   tlr 133

Qy     109 ADVDERTRKTGEAFAAGL             - APDCAITVHTQADTSSPDPLFNPLKTGV 154
           |  ::  :|| : |                  :|  ::  : :::|  :     | |  ||
Db     134 aspsdrvtasaqkflegfypqkladpatnvaaptsvsvii psetlp     atldhqv 187

Qy     155 YQLDNANVTDAILSRAGGSI ALPTSWPEVFDRQLAFRFLDIVMNFTQWAQKQDKSRLI 214
           |     |   | : :  : :|||             | |  :  : ||     :||
Db     188 v       vkttvsqplpdcvrantt             altap  dtvtvtvdkhipqvtlt 236

Qy     215 QALDSEL  KVCARNVSLDSAVSVADLFEIFSRDGWPHSAWGSIFCPWNNTDLPL 271
             |  | | |:  |   |  |  ::  :  |  |  |  |  |  |  |  |||
Db     237 dedvvsidwdwvstdtvarisdasqlqpfqplttnav       w     kkyvyfpsl 282

Qy     272 HNAQVEFLQKIEFVARSKATPLLSLIMAALEFHPEQKQ        AYGVELPTSVLFIA 323
           |:|  :  :  :  :  :  :  :  :|| | |  |  |        |  | | :
Db     283 qkyyavvtnspdapapipfttpdcdiarifvspqdtstvtfkvnpatfpldatmyy 340

Qy     324 GHDTNLANLGGALELNWTLPGQPDNTPPGGELVFEKWRRLSD-NSQWI  QVSLVFQ 377
           |  :  :  |        |  : | |       |  : |   :: |:        |:
Db     341 dtsldnsmsvsfttvl       atyntdplsttsvqvskvddpvsnswvmptarrypv 383
```

RESULT 15
AAY69556
ID AAY69556 standard; protein; 440 AA.
XX
AC AAY69556;
XX
DT 19-APR-2000 (first entry)
XX
DE Aspergillus fumigatus AD72 phytase gene.
XX
KW Phytase; myo-inositol hexakisphosphate phosphohydrolase; stabilization;
KW thermostable; animal feed; monogastric animal; phytate phosphorus;
KW phosphate availability.
XX
OS Aspergillus fumigatus AD72#2722.
XX
PN EP969089-A1.
XX
PD 05-JAN-2000.
XX
PF 23-JUN-1999; 99EP-0111464.
XX
PR 29-JUN-1998; 98EP-0111960.
XX
PA (BOFF) BOEHRINGER LA & C[HE] A.
XX
PI Brugger R; Lehmann M; Wyss M.
XX
DR WPI; 2000-094129/08.
XX
PT New stabilized enzyme formulation, useful for feed compositions for
PT monogastric animals
XX
PS Example 3; Fig 13; 101pp; English.
XX
CC The invention relates to a novel stabilised dry or liquid enzyme
CC formulation, comprising phytase (myo-inositol hexakisphosphate
CC phosphohydrolase) and one or more stabilising agents including
CC xylitol, a ribitol, polyethylene glycol with a molecular weight of 600
CC to 4000 Da, preferably 1000 to 3350 Da; the disodium salts of malonic,
CC glutaric and succinic acid; carboxymethylcellulose; and sodium alginate.
CC The stabilised phytase formulation is used in a method for preparing a
CC feed composition for monogastric animals (esp., pigs, poultry) and
CC provides a monogastric animal with its dietary requirements of
CC phosphorus. Although a large amount of phosphate is present in animal
CC feed in the form of phytate phosphorus, monogastric animals are unable
CC to utilise this form of phosphate, resulting in the addition of extra
CC phosphate to the feed of such animals. Phytase enhances the availability
CC value of plant material without the need for adding additional phosphate
CC to the feed. The level of phosphate pollution in the environment is
CC reduced by adding phytase to animal feed, as the animal can make use of
CC the inorganic phosphate liberated from phytate phosphorus using the
CC enzyme. The phytase formulation of the invention has an improved
CC thermostability and can therefore remain stable during long term storage
CC and can withstand feed processing methods such as extrusion, expansion
CC and pelleting. Sequences AAY69544 Y69556 represent the amino acid
CC sequences of mature phytases from a variety of funai (mainly different
CC strains of Aspergillus species), which were used in an exemplification of
CC the present invention to determine a consensus phytase sequence,
CC designated phytase 1 (AAY69558). An initial consensus sequence (AAY69557)
CC was determined using the program PILEUP, and the phytase 1 consensus was
CC derived from the initial consensus. Consensus phytase sequences were
CC produced in the exemplifications, as a prior art consensus phytase
CC sequence was found to have higher thermal stability than native
CC phytases.
XX
SQ Sequence 440 AA;

Query Match 5.1%; Score 118; DB 18; Length 440;
Best Local Similarity 21.7%; Prob. No. 0.0099;
Matches 102; Conservative 78; Mismatches 184; Indels 111; Gaps 11;

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: November 22, 2001, 12:05:17 ; Search time 1808.85 Seconds
(without alignments)
11313.173 Million cell updates/sec

Title: US-09-580-515-1
Perfect score: 1323
Sequence: 1 atgaaagcgatcttaatccc..........atcaccatcaccatcactaa 1323

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1341157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2608314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

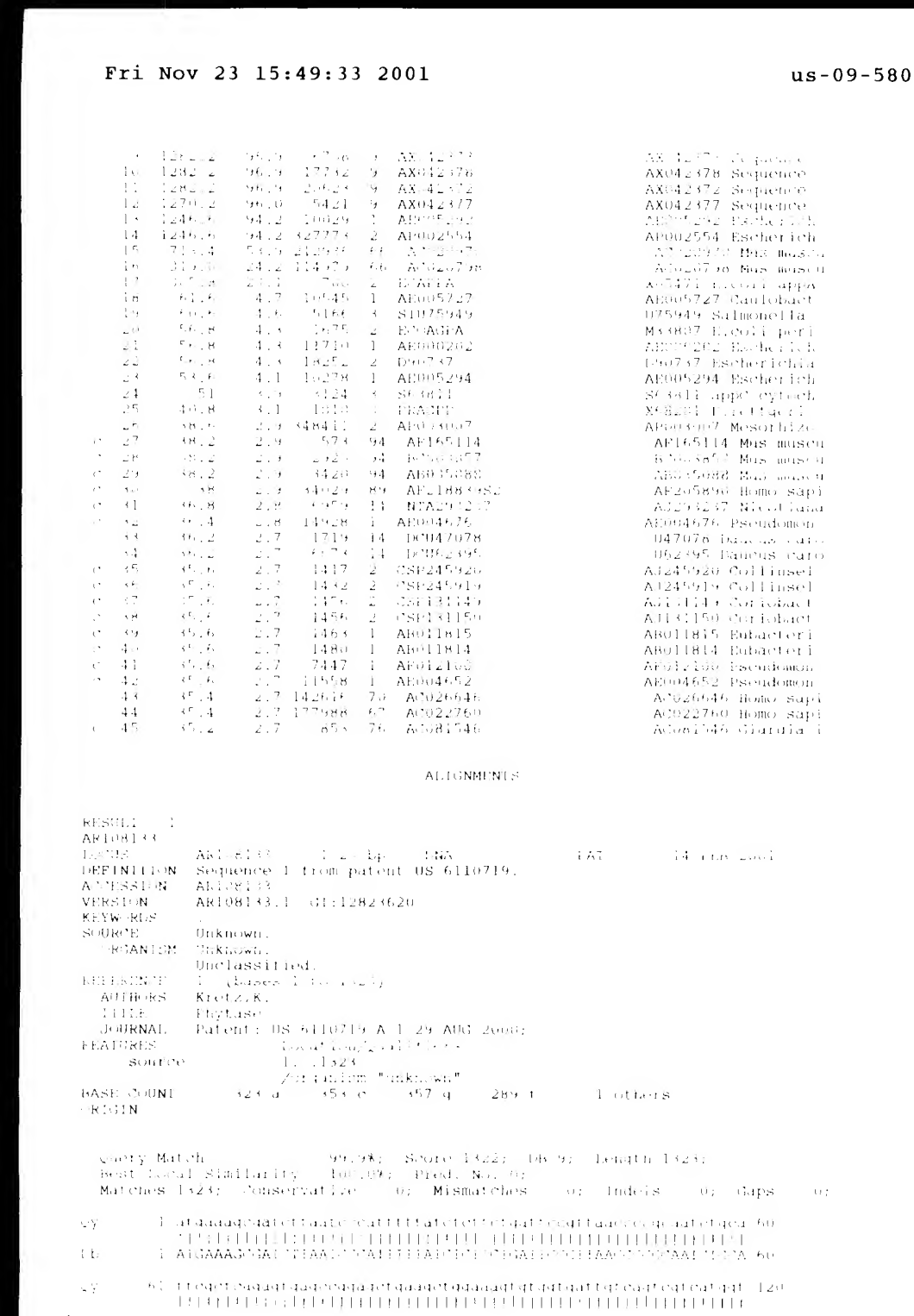
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

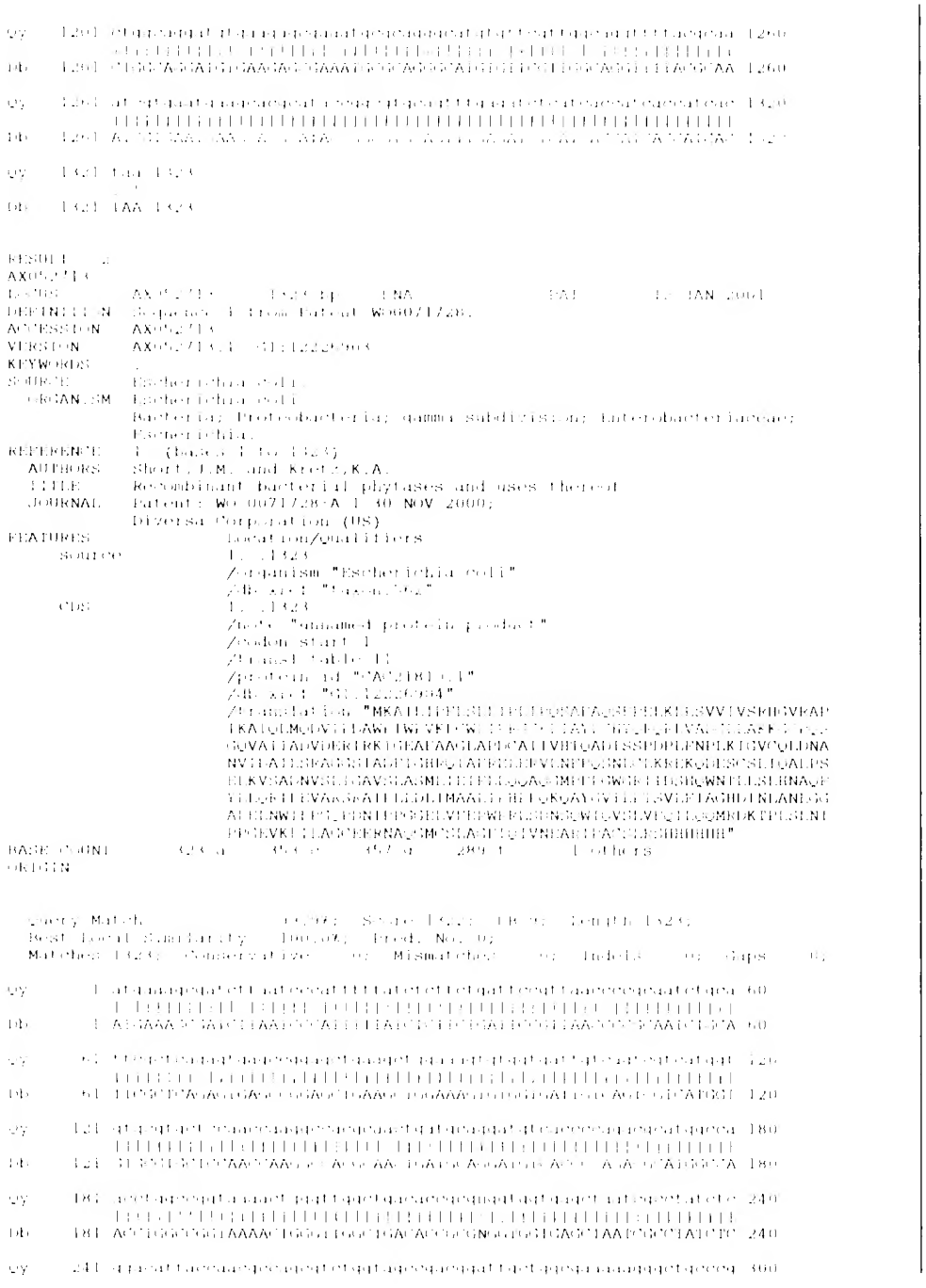
Result No.	Score	% Query Match	Length	DB	ID	Description
1	1322	99.9	1323	9	AR108133	AR108133 Sequence
2	1322	99.9	1323	9	AX052713	AX052713 Sequence
3	1287.8	97.3	1901	2	ECOALPAA	M18273 E. coli per
4	1287.8	97.3	10784	1	AE000200	AE000200 Escherich
5	1287.8	97.3	18882	2	D90735	D90735 Escherichia
6	1282.2	96.9	3470	9	AX042376	AX042376 Sequence
7	1282.2	96.9	4060	9	AX042374	AX042374 Sequence
8	1282.2	96.9	6116	9	AX042375	AX042375 Sequence



ALIGNMENTS

RESULT 1
AR108133
LOCUS AR108133 1423 bp DNA PAT 14-JUL-2001
DEFINITION Sequence 1 from patent US 6110719.
ACCESSION AR108133
VERSION AR108133.1 GI:12823620
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1423)
AUTHORS Krotz,K.
TITLE Phytase
JOURNAL Patent: US 6110719 A 1 29-AUG-2000;
FEATURES Location/Qualifiers
source 1..1423
/organism="Unknown"
BASE COUNT 423 a 453 c 457 g 289 t 1 others
ORIGIN

Query Match 99.0%; Score 1322; DB 9; Length 1423;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



RESULT 2
AX002713
LOCUS AX002713 1423 bp DNA PAT 12-JAN-2001
DEFINITION Sequence 3 from Patent WO0071728.
ACCESSION AX002713
VERSION AX002713.1 GI:12226984
KEYWORDS .
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1423)
AUTHORS Short, J.M. and Kretz, K.A.
TITLE Recombinant bacterial phytases and uses thereof
JOURNAL Patent: WO 0071728-A 1 30-NOV-2000;
Diversa Corporation (US)
FEATURES Location/Qualifiers
source 1..1423
/organism="Escherichia coli"
/db_xref="taxon:562"
CDS 1..1423
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="AAK21861.1"
/db_xref="GI:12226985"
/translation="MKAILIFFLSLLIGLLVQVAFAQSEPELKLESVVIVSRHGVRAP
TKATQLMQDVTPDAWPTWPVKLGWLTPRGEHLVALMGHYQRQKLVADGLLAKKGCPQP
GQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNA
NVTDAILSRAGGSIADFTGHRQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPS
ELKVSADNVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQF
YLRMTPTIARSRATPLLDLIKAALTPHPPQKQAYGVTLPTSVLFIAGHDTNLANLGG
ALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNT
PPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSLRGHHHHHH"
BASE COUNT 423 a 356 c 357 g 289 t 1 others
ORIGIN

Query Match 99.9%; Score 1521; DB 9; Length 1423;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db    1121 IAA 1427


RESULT   4
E.COLIAPAA
LOCUS       E.COLIAPAA      1961 bp    DNA             BCT       05-JAN-1994
DEFINITION  E. coli periplasmic phosphoanhydride phosphohydrolase (appA) gene,
            complete cds.
ACCESSION   M98708 U83370 U83371 U83372 U83373 U83374 U83375
VERSION     M98708.1  GI:145284
KEYWORDS    acid allergenol phosphatase; appA gene; periplasmic phosphoanhydride
            phosphohydrolase.
SOURCE      Escherichia coli DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1961)
  AUTHORS   Dassa,J., Marck,C. and Boquet,P.L.
  TITLE     The complete nucleotide sequence of the Escherichia coli gene appA
            reveals significant homology between pH 2.5 acid phosphatase and
            glucose-1-phosphatase
  JOURNAL   J. Bacteriol. 172 (9), 5497-5500 (1990)
  MEDLINE   90368016
REFERENCE   2  (sites)
  AUTHORS   Greenie,K., Wyrey,P.R., Stevis,P.E., Rotrel,J., Ston,M.M. and Van
            Etten,R.L.
  TITLE     Overexpression, site-directed mutagenesis, and mechanism of
            Escherichia coli acid phosphatase
  JOURNAL   J. Biol. Chem. 267 (32), 22830-22839 (1992)
  MEDLINE   93054506
REFERENCE   3  (sites)
  AUTHORS   Ostanin,K. and Van Etten,R.L.
  TITLE     Asp304 of Escherichia coli acid phosphatase is involved in leaving
            group protonation
  JOURNAL   J. Biol. Chem. 268 (28), 20778-20784 (1993)
  MEDLINE   94012612
FEATURES             Location/Qualifiers
     source          1..1961
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     RBS             3..8
                     /gene="appA"
     gene            3..1522
                     /gene="appA"
     CDS             16..1308
                     /gene="appA"
                     /codon_start=1
                     /transl_table=11
                     /product="APFX"
                     /protein_id="AAA72085.1"
                     /db_xref="GI:145284"
                     /translation="MWYLKWFVVILLMSLSTLVLVWLLFKLKS"
     stem_loop       118..149
                     /gene="appA"
     -35_signal      153..158
                     /gene="appA"
     -10_signal      156..161
                     /gene="appA"
     RBS             172..177
                     /gene="appA"
     sig_peptide     188..253
                     /gene="appA"
     CDS             188..1186
                     /gene="appA"
                     /standard_name="acid phosphatase"
                     /codon_start=1
                     /transl_table=11
                     /product="periplasmic phosphoanhydride phosphohydrolase"
                     /protein_id="AAA72086.1"
                     /db_xref="GI:145285"
                     /translation="MSAIIIFI[illegible]
                     IKAIGMGIVIIGAWFIWGVKEIWEIPPSDILAVLGVVPGPGLVAIVLLAKKGGIGQ
                     GVVLAGVLEDVTDVFWVAVIGAIFIKFIKDVDADIAIFGGLFPALLMTIFGLLFY
                     AVE                                                    
     mat_peptide     254..1483
                     /gene="appA"
                     /product="periplasmic phosphoanhydride phosphohydrolase"
     variation       299..301
                     /gene="appA"
                     /standard_name="H43A mutant"
                     /note="created by site directed mutagenesis"
                     /citation [2]
     variation       299..301
                     /gene="appA"
                     /standard_name="K92A mutant"
                     /note="created by site directed mutagenesis"
                     /citation [2]
     variation       299..301
                     /gene="appA"
                     /standard_name="K20A mutant"
                     /note="created by site directed mutagenesis"
                     /citation [2]
     variation       299..301
                     /gene="appA"
                     /standard_name="R16A mutant"
                     /note="created by site directed mutagenesis"
                     /citation [2]
     variation       302..304
                     /gene="appA"
                     /standard_name="H17N mutant"
                     /note="created by site directed mutagenesis"
                     /citation [2]
     variation       replace(1163..1165,"gaa")
                     /gene="appA"
                     /standard_name="D304E mutant"
                     /note="created by site directed mutagenesis"
                     /citation [3]
                     /phenotype="Asp 304 changed to Glu"
     variation       replace(1163..1165,"gct")
                     /gene="appA"
                     /standard_name="D304A mutant"
                     /note="created by site directed mutagenesis"
                     /citation [3]
                     /phenotype="Asp 304 changed to Ala"
     terminator      1497..1522
                     /gene="appA"
                     /note="rho independent terminator"
BASE COUNT      474 a    499 c    499 g    428 t       1 others
ORIGIN


  Query Match             97.3%;  Score 1287.8;  Db 2;  Length 1961;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 1293;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 atgaaaagcatattgatacccattttatatctctgattctgattacccttgcaaatctcgca 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160 ATGAAAAGCATATTGATACCCATTTTATATCTCTGATTCTGATTACCCTTGCAAATCTCGCA 347

Qy      61 ttagctcaaagtgaagcagtagtgtcgaaccgtgaaaatgtgcagtgattgtcaagtcgtta 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     348 TTAGCTCAAAGTGAAGCAGTAGTGTCGAACCGTGAAAATGTGCAGTGATTGTCAAGTCGTTA 507

Qy     121 gtgcctgtgcctcaaacaagtgcagtcgcacctgtagcgatgtcgcccctgtagctcctgtt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     508 GTGCCTGTGCCTCAAACAAGTGCAGTCGCACCTGTAGCGATGTCGCCCCTGTAGCTCCTGTT 667

Qy     181 acctqqccqqtaaaactqqttqctqacaccqccqaaqtqatqqaqctaatcqctatcqtc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     668 ACCTGGCCGGTAAAACTGGTTGCTGACACCGCCGAAGTGATGGAGCTAATCGCTATCGTC 827

Qy     241 qatqcaqqtaacqaaqtqqcqqtaqcqqcqqtqaaaqaaqcqqtqqcqqaaqcqqtqqcq 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     828 GATGCAGGTAACGAAGTGGCGGTAGCGGCGGTGAAAGAAGCGGTGGCGGAAGCGGTGGCG 987
```

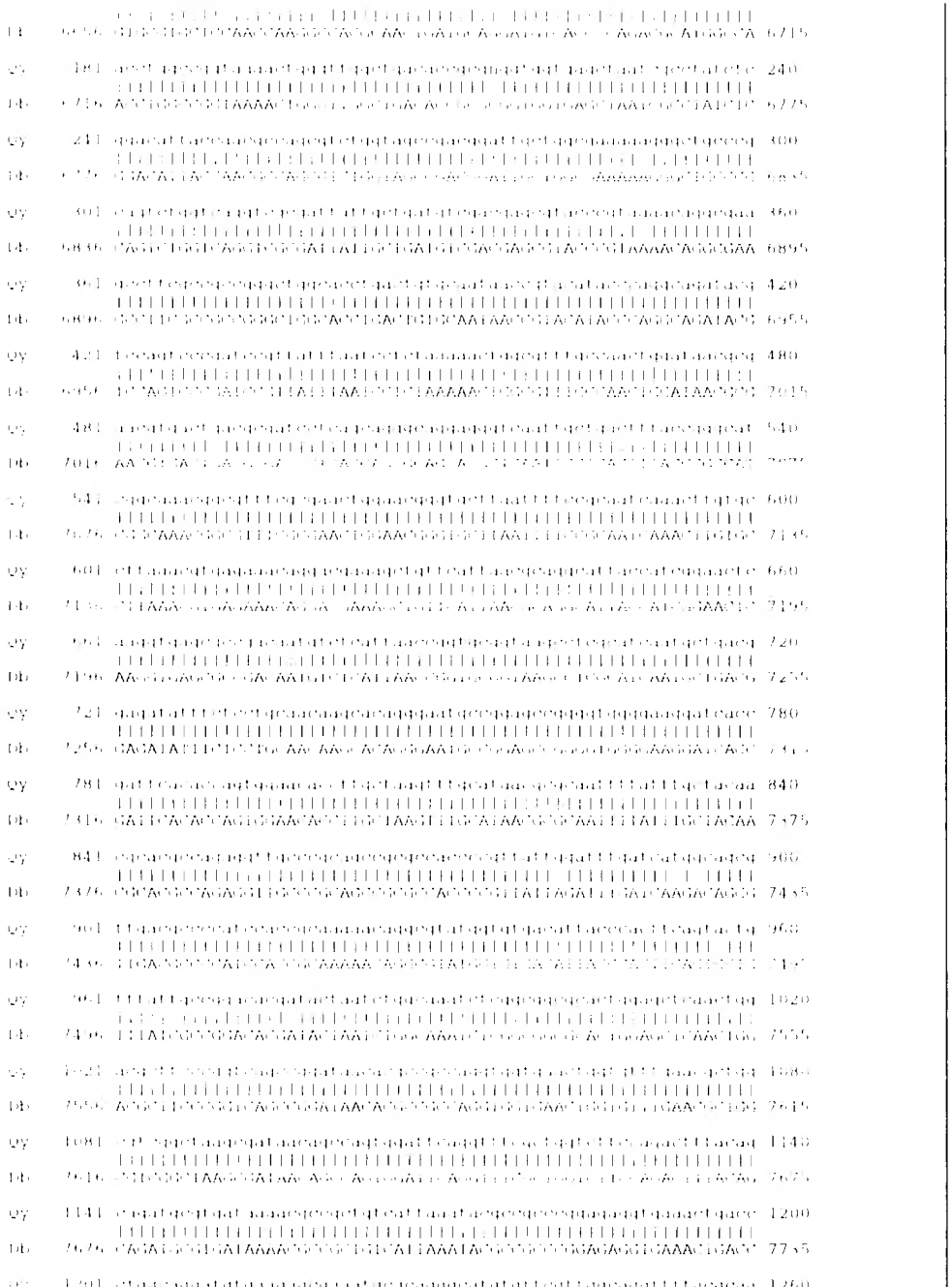
RESULT 1

AE000200
LOCUS AE000200 10784 bp DNA B'1 [illegible]
DEFINITION Escherichia coli K12 MG1655 section 90 of 400 of the complete genome.
ACCESSION AE000200 U00096
VERSION AE000200.1 GI:2367111
KEYWORDS .
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10784)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12.
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 10784)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16 JAN 1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608 262 2594 Fax: 608 263 7459
REFERENCE 3 (bases 1 to 10784)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02 SEP 1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608 262 2594 Fax: 608 263 7459
REFERENCE 4 (bases 1 to 10784)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (15-OCT 1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT On Sep 9, 1998 this sequence version replaced gi:2367211
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants RG00301 and RG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@amber.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CGSC Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein or RNA encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES Location/Qualifiers
source 1..10784
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"

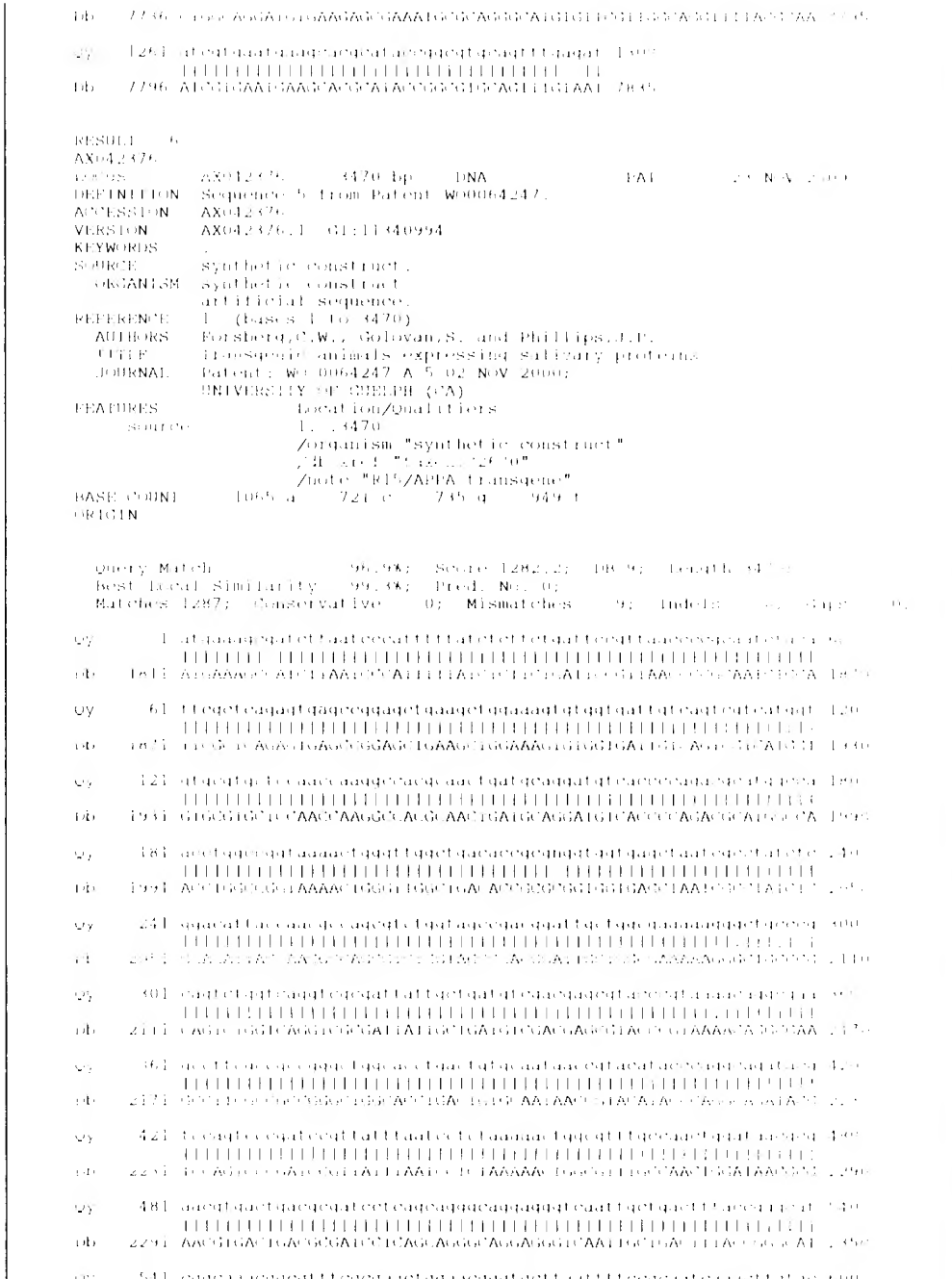
```
repeat_region   49..84
                /note="REP (repetitive extragenic palindromic) element;
                contains 1 REP sequence"
gene            123..1421
                /gene="appA"
                /note="b0980"
CDS             123..1421
                /gene="appA"
                /EC_number="3.1.3.2"
                /EC_number="3.1.3.26"
                /function="enzyme; Central intermediary metabolism: Pool,
                multipurpose conversions"
                /note="o432; 100 pct identical to PPA_ECOLI SW: P07102"
                /codon_start=1
                /transl_table=11
                /product="phosphoanhydride phosphorylase; pH 2.5 acid
                phosphatase; periplasmic"
                /protein_id="AAC74065.1"
                /db_xref="GI:1787215"
                /translation="MKAILIPFLSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAP
                [illegible]
                NVIDAILSRAGGSIADFTGHRQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPS
                [illegible]
                [illegible]
                [illegible]
                [illegible]"
repeat_region   1153..1189
                /note="REP (repetitive extragenic palindromic) element;
                contains 1 REP sequence"
repeat_region   1491..1524
                /note="REP (repetitive extragenic palindromic) element;
                contains 1 REP sequence"
gene            complement(1536..3716)
                /gene="yccC"
                /note="b0981"
CDS             complement(1536..3716)
                /gene="yccC"
                /function="orf; [illegible]"
                /note="f726; 100 pct identical to fragment [illegible]
                SW:P38134 but has 627 additional N-terminal aa and 2 less
                C-terminal residues"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC74066.1"
                /db_xref="GI:1787216"
                /translation="[illegible]
                [illegible]
                [illegible]
                GHYILEGEEFIVRGMVGQKLEKGVALIIAEIFAFE[illegible]
                [illegible]
                PEVKGEEGQAHEKLNVYKQLE[illegible]
                [illegible]
                [illegible]
                RGVEPF[illegible]
                LSAVEAVEALFIGLHFMMEIENNIEM[illegible]
                DADLRKGYGRNEIVVNNEHNFPYIAQKFELNEVIQHPKKTTVIIEKQVFFNPGHL
                LMEIRMRQLLEWANEHYDLVIVDTPPMLAVSDAAVVGRVGTSLLVARFGENTAKEVS
                LSMQRLEQAGVNIKGAILNDVIKRASAYATSGYNTYSPYEKE"
promoter        complement(3733..3764)
                /note="factor Sigma70; predicted +1 start at 104344"
gene            complement(3736..4194)
                /gene="yccY"
                /note="b0982"
CDS             complement(3736..4194)
                /gene="yccY"
                /function="putative enzyme; Not classified"
                /note="f152; This 152 aa ORF is 33 pct identical (13 gaps)
                to 136 residues of an approx. 160 aa protein PTPA_SCHPO
                SW: P41893"
                /codon_start=1
                /transl_table=11
                /product="putative phosphatase"
                /protein_id="AAC74067.1"
                /db_xref="GI:1787217"
                /translation="[illegible]
                [illegible]
                [illegible]"
promoter        complement(3793..3820)
                /gene="yccY"
                /note="factor Sigma70; predicted +1 start at [illegible]"
promoter        complement(3818..3844)
                /gene="yccY"
                /note="factor Sigma70; predicted +1 start at [illegible]"
promoter        complement(3865..3892)
                /gene="yccY"
                /note="factor Sigma70; predicted +1 start at [illegible]"
gene            complement(4170..5309)
                /gene="yccZ"
                /note="b0983"
CDS             complement(4170..5309)
                /gene="yccZ"
                /function="phenotype; Not classified"
                /note="f379; This 379 aa ORF is 29 pct identical (55 gaps)
                to 364 residues of an approx. 460 aa protein [illegible]
                SW: P22236"
                /codon_start=1
                /transl_table=11
                /product="putative function in exopolysaccharide
                production"
                /protein_id="AAC74068.1"
                /db_xref="GI:1787218"
                /translation="MKENIFKPSVLLLAVLSLIA[illegible]
                [illegible]
                [illegible]
                VDVSVAAPRSQKAYVT[illegible]
                [illegible]
                [illegible]
                EFYLIVYVIIAPLARWNRVIALLVPFI[illegible]"
gene            complement(5455..7451)
                /gene="ymcA"
                /note="b0984"
CDS             complement(5455..7451)
                /gene="ymcA"
                /function="orf; Unknown"
                /note="f698; This 698 aa ORF is 64 pct identical (7 gaps)
                to 694 residues of an approx. 704 aa protein [illegible]
                SW: P32689"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC74069.1"
                /db_xref="GI:1787219"
                /translation="[illegible]
                [illegible]
                [illegible]
                [illegible]"

Query Match             97.3%;  Score 1287.8;  DB 1;  Length 10234;
Best Local Similarity   99.4%;  Pred. No. 0;
Matches 1292;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 atgaaagcgatcttaatcccatttttatctcttctgattccgttaacc[illegible]
Db        123 ATGAAAGCGATCTTAATCCCATTTTTATCTCTTCTGATTCCGTTAACC[illegible]

Qy         61 [illegible]
Db        183 [illegible] 242

Qy        121 [illegible] 180
Db        243 [illegible]

Qy        181 [illegible]
```

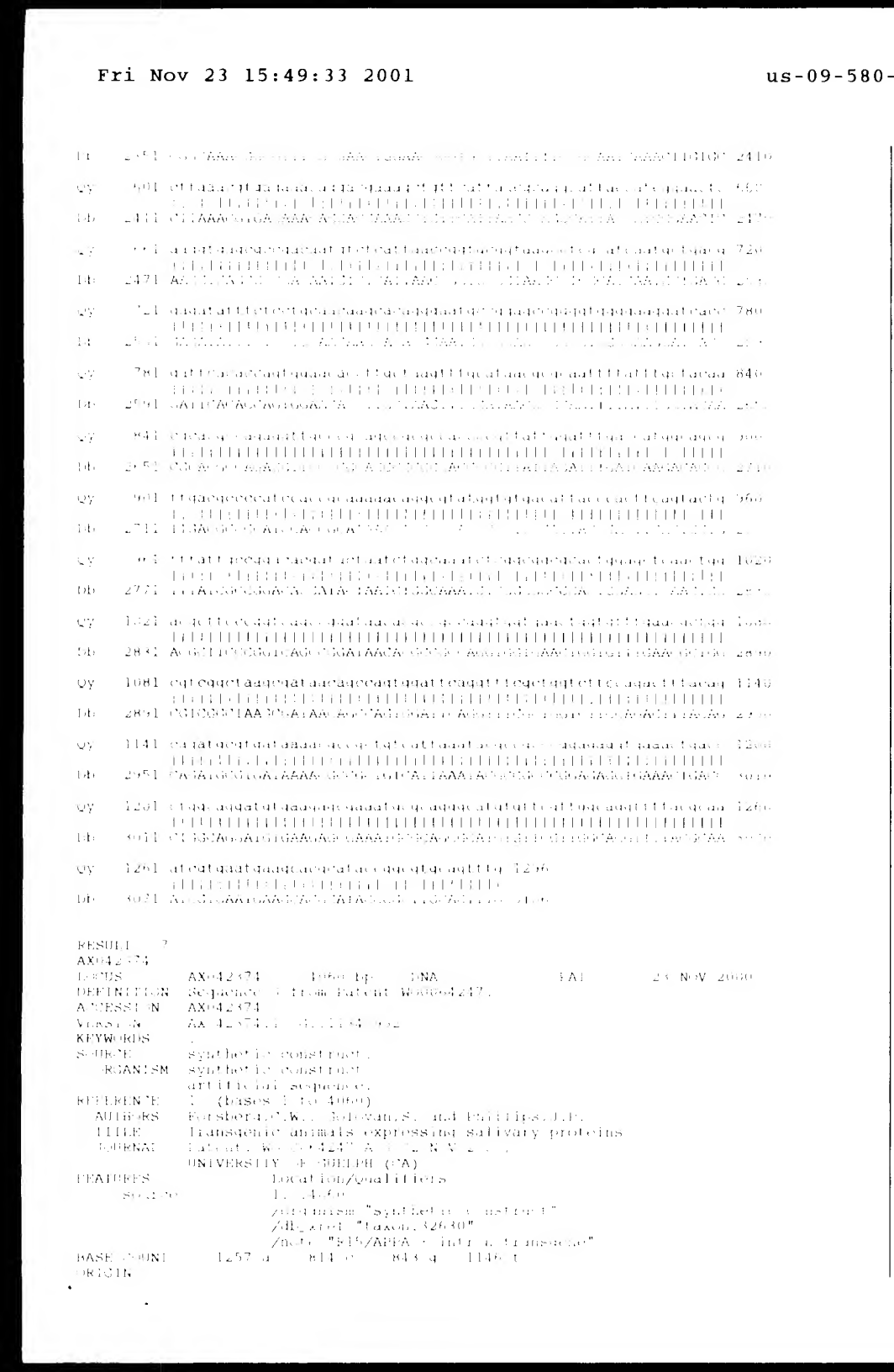
```
Db        303 [illegible] 362

Qy        241 [illegible] 300
Db        363 [illegible] 422

Qy        301 [illegible] 360
Db        423 [illegible] 482

Qy        361 [illegible] 420
Db        483 [illegible] 542

Qy        421 [illegible] 480
Db        543 [illegible] 602

Qy        481 [illegible] 540
Db        603 AACTGACTGACGCGATCCTCAGCAGGCAGGAGGTCAATTGCTGACTTTACCGGGCAT 662

Qy        541 [illegible] 600
Db        663 [illegible] 722

Qy        601 [illegible] 660
Db        723 [illegible] 782

Qy        661 [illegible] 720
Db        783 [illegible] 842

Qy        721 [illegible] 780
Db        843 [illegible] 902

Qy        781 [illegible] 840
Db        903 [illegible] 962

Qy        841 [illegible] 900
Db        963 [illegible] 1022

Qy        901 [illegible] 960
Db       1023 [illegible] 1082

Qy        961 [illegible] 1020
Db       1083 [illegible] 1142

Qy       1021 [illegible] 1080
Db       1143 [illegible] 1202

Qy       1081 [illegible] 1140
Db       1203 [illegible] 1262

Qy       1141 [illegible] 1200
Db       1263 [illegible] 1322

Qy       1201 [illegible] 1260
Db       1323 [illegible] 1382

Qy       1261 [illegible] 1300

Db       1383 [illegible] 1422


RESULT      5
D90735
LOCUS       D90735     18882 bp    DNA             BCT       02-FEB-1999
DEFINITION  Escherichia coli genomic DNA. (22.3 - 22.7 min).
ACCESSION   D90735 AB001340
VERSION     D90735.1  GI:4062539
KEYWORDS    Complete and shotgun sequencing; hyaB; hyaC; hyaD; [illegible]
            appC; cyxA; cbdA; appB; cyxB; cbdB; appA; [illegible]
            espB; sfa; yccL; yccM.
SOURCE      Escherichia coli(strain:K12) DNA, clone:Kohara clone #[illegible]
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
REFERENCE   2  (sites)
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., [illegible]
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 18882)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1996) to the DDBJ/EMBL/GenBank databases.
            Hirotada Mori, NARA Institute of Science and Technology, Res. &
            Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01,
            Japan (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-[illegible],
            Fax:81-7437-2-5669)
COMMENT     On Dec 24, 1998 this sequence version replaced gi:[illegible].
            Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                      Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http://ecoli.aist-nara.ac.jp.
FEATURES             Location/Qualifiers
```

```
source          1..18882
                /organism "Escherichia coli"
                /strain "K12"
                /db_xref "taxon:562"
                /clone "Kohara clone #225"
                /map "22.3-22.7 min"
                /note "Nucleotide position [illegible] from the
                initiation site of thrA (0 min.). This clone is from
                Kohara lambda miniset library."
gene            1..[illegible]
                /gene "hyaB"
CDS             [illegible]
                /gene "hyaB"
                /note "ORF_ID:o225#[illegible]
                similar to SwissProt Accession Number [illegible]"
                /codon_start 1
                /transl_table 11
                /product "Hydrogenase 1 large chain (EC 1.18.99.1) (NiFE
                hydrogenase) (membrane-bound hydrogenase 1 large subunit)
                (hyaB)."
                /protein_id "BAA35738.1"
                /db_xref "GI:4062540"
                /translation "[illegible]
                MPRGAVINGDFNNVIPVDLVDPQQVQIFVDHAWYRYPNEQVGKHPFDGITDPWYNPGD
                [illegible]
                [illegible]
                [illegible]
                MNIKMAIIPQFIEILRILHSFDPCLACSIHVLGDGGSEIISVQVK"
gene            985..1[illegible]
                /gene "hyaC"
CDS             985..1[illegible]
                /gene "hyaC"
                /note "ORF_ID:o225#1
                similar to PIR Accession Number [illegible]"
                /codon_start 1
                /transl_table 11
                /product "[illegible]"
                /protein_id "BAA[illegible]"
                /db_xref "GI:1651475"
                /translation "[illegible]
                [illegible]
                [illegible]
                [illegible]
                NCYRSHKFCKISNKERS"
gene            1689..2276
                /gene "hyaD"
CDS             1689..227[illegible]
                /gene "hyaD"
                /note "ORF_ID:o225#2
                similar to SwissProt Accession Number [illegible]"
                /codon_start 1
                /transl_table 11
                /product "Hydrogenase-1 operon protein hyaD."
                /protein_id "BAA[illegible]"
                /db_xref "GI:1651476"
                /translation "[illegible]
                [illegible]
                [illegible]
                [illegible]"
gene            227[illegible]..2671
                /gene "hyaE"
CDS             227[illegible]..2671
                /gene "hyaE"
                /note "ORF_ID:o225#3
                similar to SwissProt Accession Number [illegible]"
                /codon_start 1
                /transl_table 11
                /product "Hydrogenase-1 operon protein hyaE."
                /protein_id "BAA35741.1"
                /db_xref "GI:1651477"
                /translation "[illegible]
                [illegible]
                YRCVLNGIHPWALLINLMRGIVDEQQHRAQ"
gene            2668..3[illegible]25
                /gene "hyaF"
CDS             2668..3[illegible]25
                /gene "hyaF"
                /note "ORF_ID:o225#4
                similar to SwissProt Accession Number [illegible]"
                /codon_start 1
                /transl_table 11
                /product "Hydrogenase-1 operon protein hyaF."
                /protein_id "BAA35742.1"
                /db_xref "GI:1651478"
                /translation "[illegible]
                [illegible]
                [illegible]
                [illegible]
                [illegible]
                QWLAEAAPI"
gene            3659..5203
                /gene "appC"
                /note "cbdA; cyxA"
CDS             3659..5203
                /gene "appC"
                /note "ORF_ID:o225#5
                similar to SwissProt Accession Number P26459"
                /codon_start 1
                /transl_table 11
                /product "Cytochrome bd II oxidase subunit I (EC
                1.10.3.-)."
                /protein_id "BAA35743.1"
                /db_xref "GI:1651479"
                /translation "[illegible]
                [illegible]
                [illegible]
                [illegible]
                [illegible]
                [illegible]
                [illegible]
                [illegible]
                [illegible]
                PIQQQG"
gene            5215..6351
                /gene "appB"
                /note "cbdB; cyxB"
CDS             5215..6351
                /gene "appB"
                /note "ORF_ID:o225#6
                similar to SwissProt Accession Number [illegible]"
                /codon_start 1
                /transl_table 11
                /product "Cytochrome bd II oxidase subunit II (EC
                1.10.3.-)."
                /protein_id "BAA35744.1"
                /db_xref "GI:1651480"
                /translation "[illegible]
                [illegible]
                [illegible]
                [illegible]
                CTVLWVCIGCVLLAQGANGISNIGMKLVAVLPCAWHNNFVESPVLWIFPLLGFFCPL
                LLVMAIYRGM[illegible]"

  Query Match             97.7%;  Score 1287.8;  DB 2;  Length 18882;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1292;  Conservative    0;  Mismatches    8;  Indels    [illegible];  Gaps    [illegible];

Qy          1 atgaaaqgqatcttaatcccattttttatcttttctqattccqttaacccc[illegible]
              [illegible]
Db       [illegible] ATGAAAG[illegible] 6595

Qy         61 ttcgqtcaqqqtqaqccqqaqctqaaqctqqaaaqtqtqqtqatt[illegible] 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       [illegible]

Qy        121 qtqcqtqctccaaccaaqqccacqcaactqatqcaqqatqtcaccccaqacqcatqq[illegible] 180
```

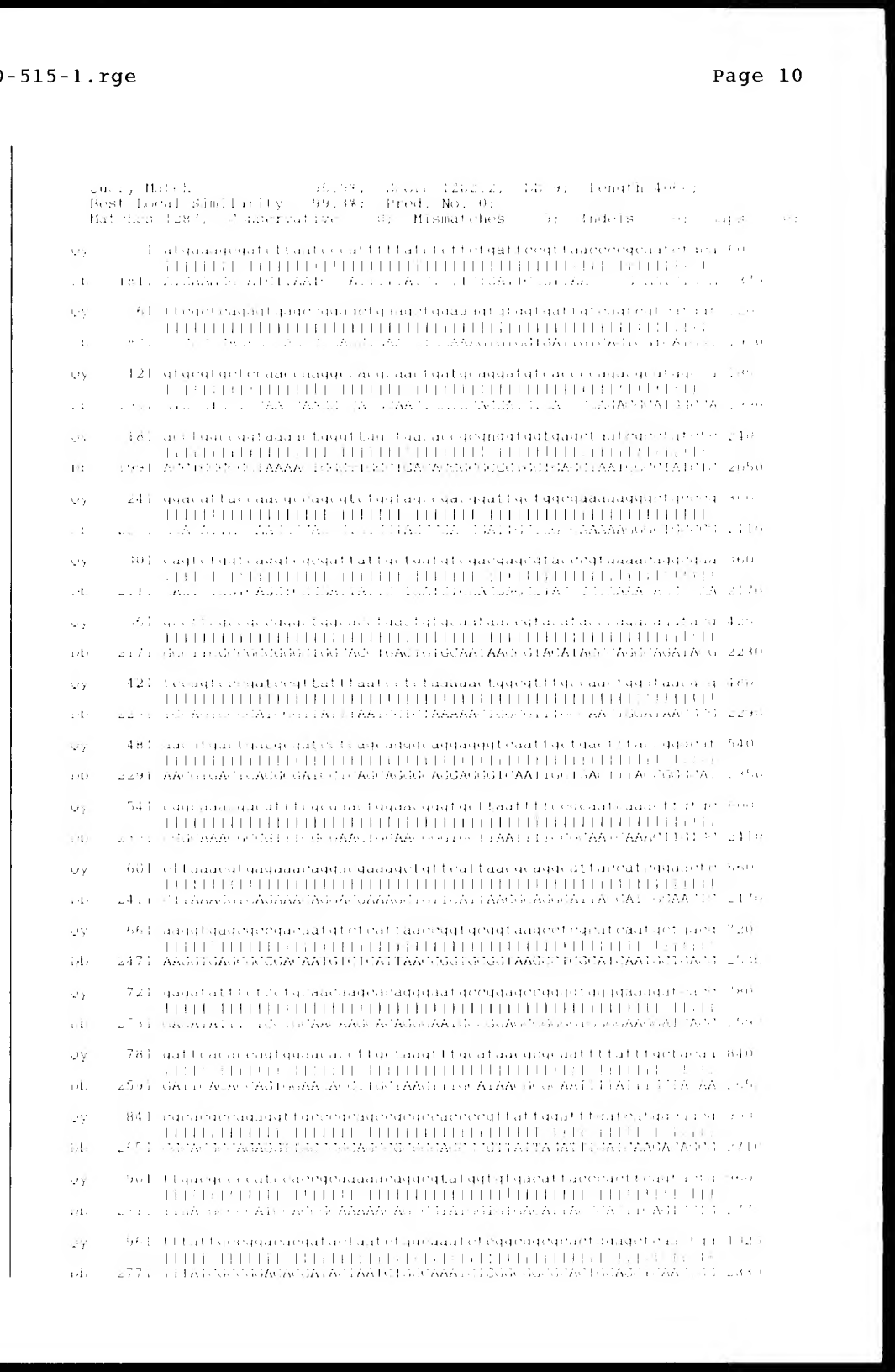


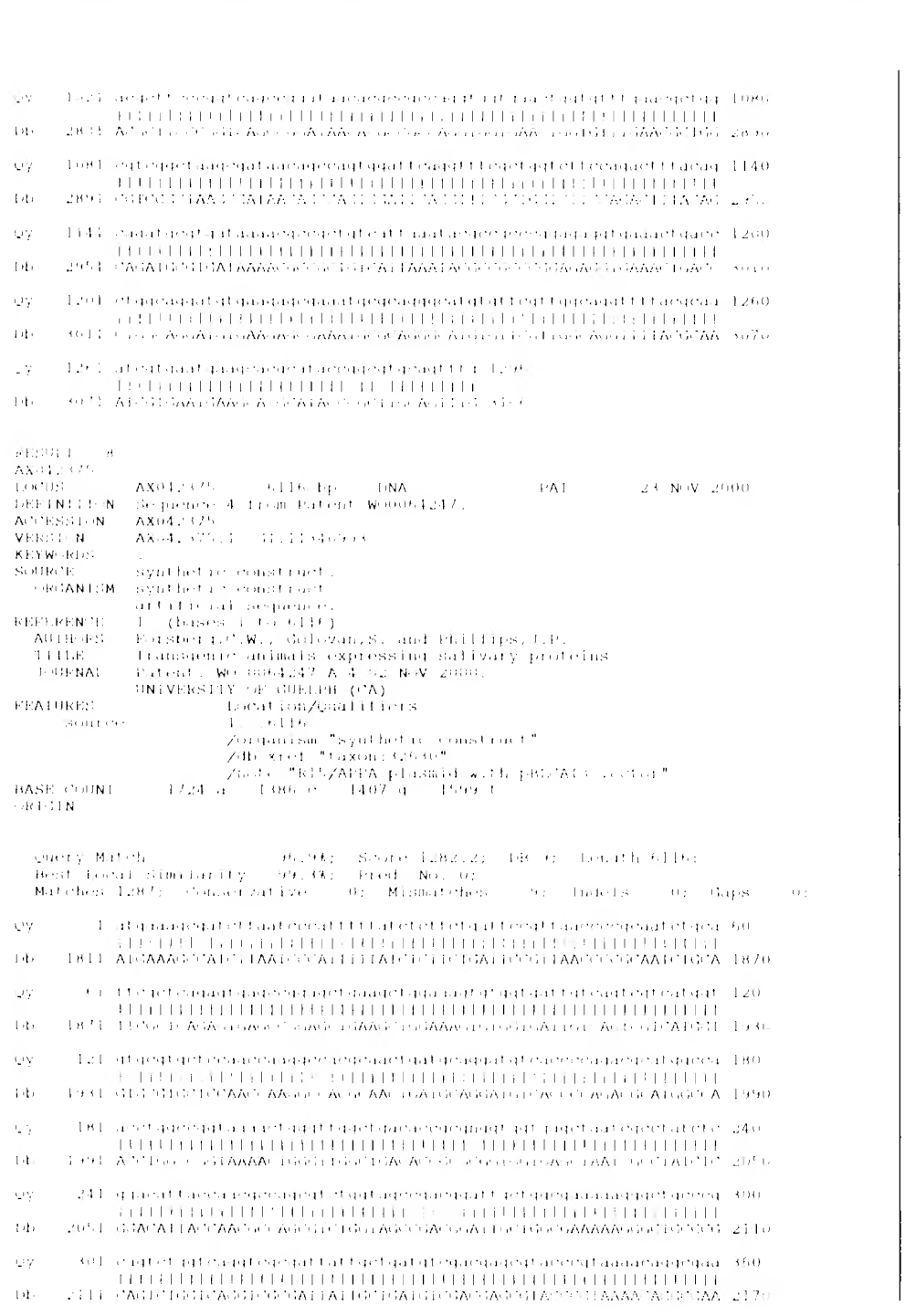
```
RESULT   6
AX042376
LOCUS       AX042376       3470 bp    DNA             PAT       25-NOV-2000
DEFINITION  Sequence 6 from Patent WO0064247.
ACCESSION   AX042376
VERSION     AX042376.1  GI:11340994
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 3470)
  AUTHORS   Forsberg,C.W., Golovan,S. and Phillips,J.P.
  TITLE     Transgenic animals expressing salivary proteins
  JOURNAL   Patent: WO 0064247 A 5 02-NOV-2000;
            UNIVERSITY OF GUELPH (CA)
FEATURES             Location/Qualifiers
     source          1..3470
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /note="R15/APPA transgene"
BASE COUNT      1065 a    721 c    735 g    949 t
ORIGIN
```





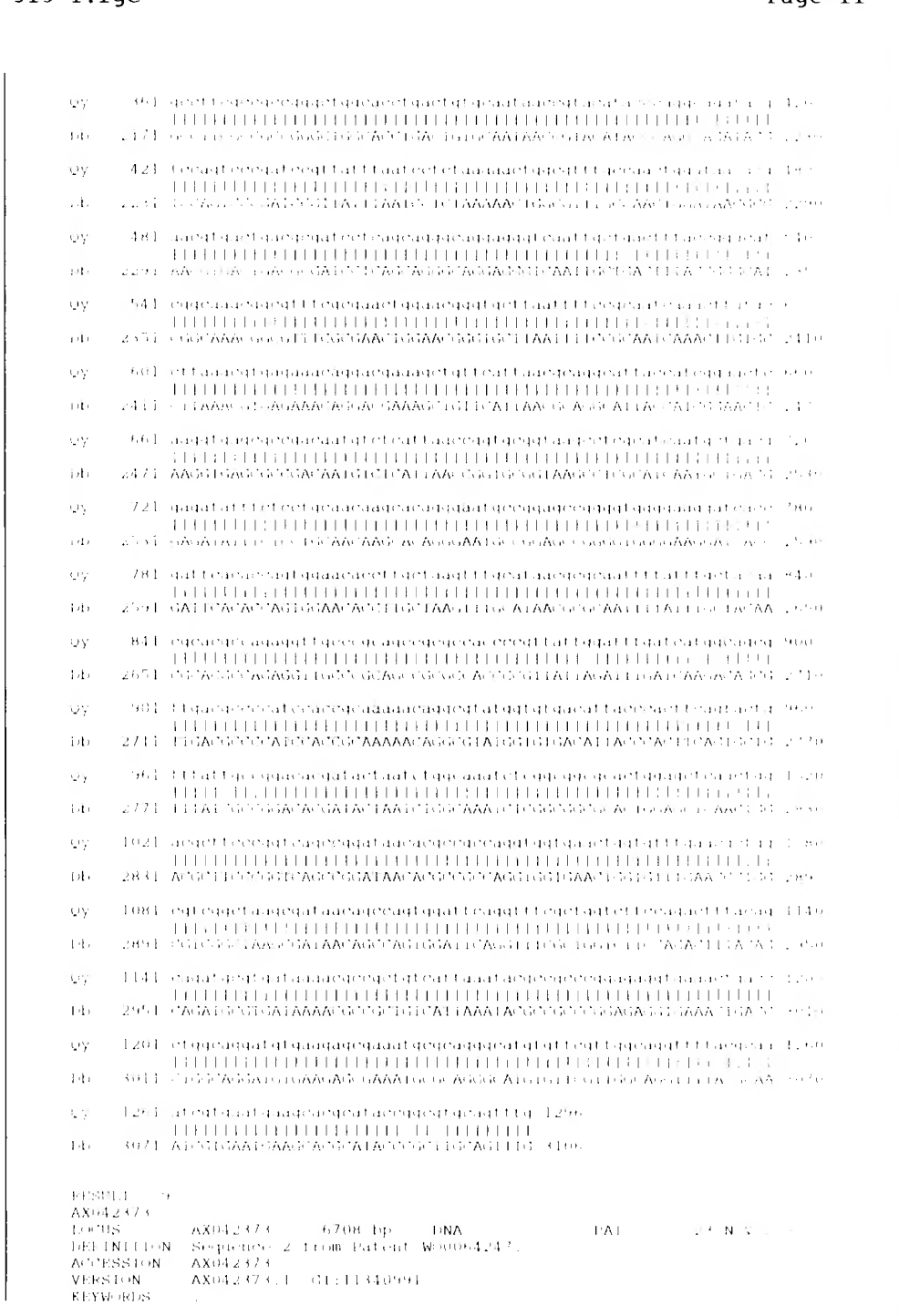
```
RESULT   7
AX042374
LOCUS       AX042374      1060 bp    DNA             PAT       23-NOV-2000
DEFINITION  Sequence 7 from Patent WO0064247.
ACCESSION   AX042374
VERSION     AX042374.1  GI:13429902
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Forsberg,C.W., Solomon,S. and Phillips,J.P.
  TITLE     Transgenic animals expressing salivary proteins
  JOURNAL   Patent: WO 0064247-A 1 26-OCT-2000;
            UNIVERSITY OF GUELPH (CA)
FEATURES             Location/Qualifiers
     source          1..1060
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="SIV/APPA - Intron transgene"
BASE COUNT     257 a    314 c    343 g    146 t
ORIGIN
```





LOCUS AX042373 6116 bp DNA PAT 23-NOV-2000
DEFINITION Sequence 4 from Patent WO0064247.
ACCESSION AX042373
VERSION AX042373.1 GI:11340993
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 6116)
AUTHORS Forster,J.C.W., Golovan,S. and Phillips,J.P.
TITLE Transgenic animals expressing salivary proteins
JOURNAL Patent: WO 0064247 A 3 02-NOV-2000;
UNIVERSITY OF GUELPH (CA)
FEATURES Location/Qualifiers
source 1..6116
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="R15/AEPA plasmid with pMGTA1 vector"
BASE COUNT 1724 a 1886 c 1407 g 1099 t
ORIGIN

Query Match 96.9%; Score 1282.2; DB 0; Length 6116;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



LOCUS AX042373 6708 bp DNA PAT 23-NOV-2000
DEFINITION Sequence 2 from Patent WO0064247.
ACCESSION AX042373
VERSION AX042373.1 GI:11340991
KEYWORDS .

```
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 6708)
  AUTHORS   Forsberg,C.W., Golovan,S. and Phillips,J.P.
  TITLE     Transgenic animals expressing salivary proteins
  JOURNAL   Patent: WO 0064247 A 7 02 NOV 2000;
            UNIVERSITY OF GUELPH (CA)
FEATURES             Location/Qualifiers
     source          1..6708
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="bAP/AMPA transgene with pBS/AMY vector"
BASE COUNT     1416 a    1475 c    1515 g    1302 t
ORIGIN
```

Query Match 96.9%; Score 1282.2; DB 9; Length 6708;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy       1 atgaaaagcatattaatcccattttatatctttctgattttccttaaccccctcaatctcca   60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1871 ATGAAATTTCTCTTAATCCCATTTTATATCTTTCTGATTTTCCTTAACCCCCTCAATCTCCA 1930

Qy      61 ttcgctcagaagtgagcgaacgcagctgaactgaaaatgtgatgatgattgctgactctatgt   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1931 TTCGCTCAGAAGTGAGCGAACGCAGCTGAACTGAAAATGTGATGATGATTGCTGACTCTATGT 1990

Qy     121 atacgtactccaacaaaacaacgaacttattcgacgaatgcaccccaacacgatgacgtcca   180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1991 GTACGTACTCCAACAAAACAACGAACTTATTCGACGAATGCACCCCAACACGATGACGTCCA 2050

Qy     181 agctaacgcagtaaaactgattgggctgacgccgtcgatgaagtgtgacttcgccgatatctc   240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2051 AGCTAACGCAGTAAAACTGATTGGGCTGACGCCGTCGATGAAGTGTGACTTCGCCGATATCTC 2110

Qy     241 cgaacattacgaaacgccagcatctgtatgacgcgaacgattgctgtcgctaaaaaagatggtg   300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2111 CGAACATTACGAAACGCCAGCATCTGTATGACGCGAACGATTGCTGTCGCTAAAAAGATGGTG 2170

Qy     301 gacgttaacgtcgcagtcactgaagaaactgaagcgctggatgcgggcgaagcagacgtgacg   360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2171 GACGTTAACGTCGCAGTCACTGAAGAAACTGAAGCGCTGGATGCGGGCGAAGCAGACGTGACG 2230

Qy     361 gacttcgccgccgcaactgcatttatgatcctcaaaacttcgatgacgaagacgaagacgag   420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2231 GACTTCGCCGCCGCAACTGCATTTATGATCCTCAAAACTTCGATGACGAAGACGAAGACGAG 2290

Qy     421 taagatccgatcattattaatcctaaaaatttcgttgtcgtcgaaaatgaagatgaagtcgt   480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2291 TAAGATCCGATCATTATTAATCCTAAAAATTTCGTTGTCGTCGAAAATGAAGATGAAGTCGT 2350

Qy     481 aacgtgactgacgcgatcctcagcgagggcaagaagatcgattgctgactttgccgagcgcat   540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2291 AACGTGACTGACGCGATCCTCAGCGAGGGCAAGAAGATCGATTGCTGACTTTGCCGAGCGCAT 2350

Qy     541 cgacaaaccgcgttccgcgaacgtgaaacggctgttaatttccgcaatcaaactgtatgcc   600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2351 CGACAAACCGCGTTCCGCGAACGTGAAACGGCTGTTAATTTCCGCAATCAAACTGTATGCC 2410

Qy     601 ttaaacgtaatcaacaacgaaatgaagctgttcattgacgccgaacgtatcgatgaatacgaactc   660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2411 CTTAAACGTATCAACAACGAAATGAAGCTGTTCATTGACGCCGAACGTATCGATGAATACGAACTC 2470

Qy     661 aagatgaagcgcgaagatatgtctgattacgatgcgtatgaagctgaacgcatcgatgaacgc   720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2471 AAGATGAAGCGCGAAGATATGTCTGATTACGATGCGTATGAAGCTGAACGCATCGATGAACGC 2530

Qy     721 cgagatattctgaatctcaacgagcgcgaagtgattcgcgacgcgctgaacgatgatcagcgc   780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2531 CGAGATATTCTGAATCTCAACGAGCGCGAAGTGATTCGCGACGCGCTGAACGATGATCAGCGC 2590
```

```
Qy     781 gattcgaacgaagtgatgaacgatttgattgatttgcatgaagcgcgcaatttgattgatct   840
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2591 GATTCGATCGATCATGCGAACATCTTGCTAAGTTTGCATAACGCGCGAATTTGATT--ATCT 2648

Qy     841 cgcaacgccaagagtttgccgcgcatcgcgcgcccgcgttattgaatttgatgcatgcgcgct   900
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||| | |||||
Db    2649 CGCAACGCCAAGAGTTTGCCGCGCATCGCGCGCCGCGTTATTGAATTTGATGCATGCGCGCT 2708

Qy     901 ttgaacgcgcgaatcgaacgcgcaagaacagcgtgatgatgatgaacgatttaccgcattgatc   960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2709 TTGAACGCGCGAATCGAACGCGCAAGAACAGCGTGATGATGATGAACGATTTACCGCATTGATC 2768

Qy     961 ttgatttcgaagaagatgactatcgagaagatctcgcgcgtgcgcaatttgatttgcaacgct   1020
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2769 TTGATTTCGAAGAAGATGACTATCGAGAAGATCTCGCGCGTGCGCAATTTGATTTGCAACGCT 2828

Qy    1021 acgactacgatgacgacgatgacgacgaagatgatgaacgtgatgatttgattgatgtgatcaac   1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2829 ACGACTACGATGACGACGATGACGACGAAGATGATGAACGTGATGATTTGATTGATGTGATCAAC 2888

Qy    1081 cgtcgagctaagcgaatgaacgcgatgatgaacgcgatgatgaacgcgcgcgcgatgatgaacgct   1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2889 CGTCGGCTAAGCGAATGAACGCGATGATGAACGCGATGATGAACGCGCGCGCGATGATGAACGCT 2948

Qy    1141 cagatgcatgatgcaaaagcgcgcgatgatgaacgcgcgcgcgatgatgaacgctacgcgcgc   1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2949 CAGATGCATGATGCAAAAGCGCGCGATGATGAACGCGCGCGCGATGATGAACGCTACGCGCGC 3008

Qy    1201 gtgagcgaagtgatgaacgaagatgcgcagcaagcgcatgatgatgcgtggcgtggtgcgcgc   1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3009 GTGAGCGAAGTGATGAACGAAGATGCGCAGCAAGCGCATGATGATGCGTGGCGTGGTGCGCGC 3068

Qy    1261 atcgtgaatgaagcgcgcgatacgcgcgatgcgtgaatttgt 1296
           ||||||||||||||||||||||||||||||||||||||||||| ||
Db    3069 ATCGTGAATGAAGCGCGCGATACGCGCGATGCGTGAATTTGTT 3110
```

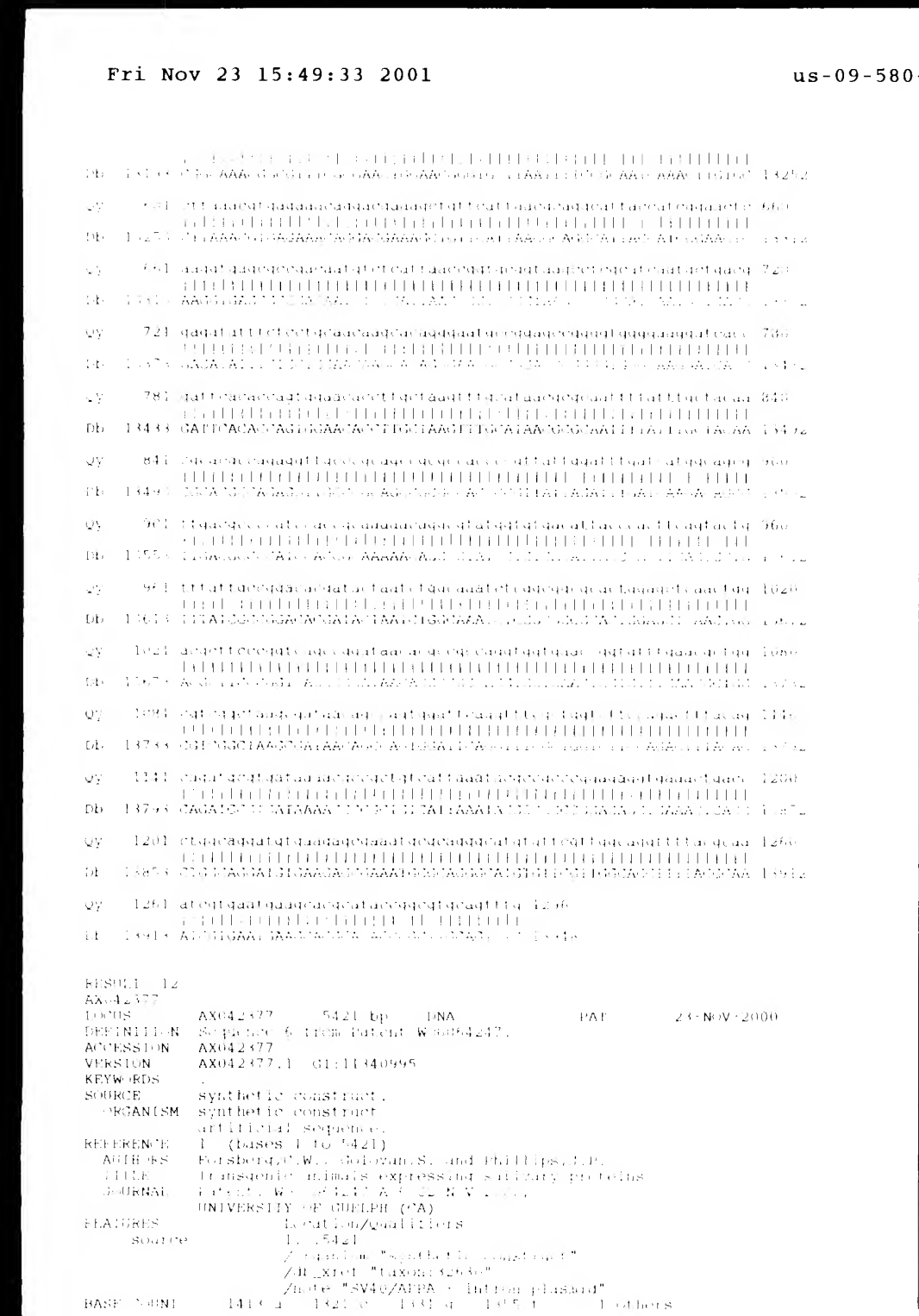
```
RESULT  10
AX042378
...e...     GS_Lu...  ...Ju...  ...bc                 rev         ... N...
DEFINITION  Sequence 7 from Patent WO0064247.
ACCESSION   AX042378
VERSION     AX042378.1  GI:11340996
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 17742)
  AUTHORS   Forsberg,C.W., Golovan,S. and Phillips,J.P.
  TITLE     Transgenic animals expressing salivary proteins
  JOURNAL   Patent: WO 0064247 A 7 02 NOV 2000;
            UNIVERSITY OF GUELPH (CA)
FEATURES             Location/Qualifiers
     source          1..17742
                     /organism="synthetic construct"
                     /db_xref="taxon:32640"
                     /note="Lama2/AMPA transgene"
BASE COUNT     4719 a    4125 c    4168 g    4729 t       1 others
ORIGIN
```

Query Match 96.9%; Score 1282.2; DB 9; Length 17742;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy       1 atgaaaagcatattaatcccattttatatctttctgattttccttaaccccctcaatctcca   60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12712 ATGAAATTTCTCTTAATCCCATTTTATATCTTTCTGATTTTCCTTAACCCCCTCAATCTCCA 12771

Qy      61 ttcgctcagaagtgagcgaacgcagctgaactgaaaatgtgatgatgattgctgactctatgt   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12772 TTCGCTCAGAAGTGAGCGAACGCAGCTGAACTGAAAATGTGATGATGATTGCTGACTCTATGT 12831
```

RESULT 11
AX042372
LOCUS AX042372 20623 bp DNA PAT 25-NOV-
DEFINITION Sequence 1 from Patent WO0064247.
ACCESSION AX042372
VERSION AX042372.1 GI:11340990
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 20623)
AUTHORS Forsberg,C.W., Goloravan,S. and Phillips,J.P.
TITLE Transgenic animals expressing salivary proteins
JOURNAL Patent: WO 0064247-A 1 02-NOV 2000;
University of Guelph (CA)
FEATURES Location/Qualifiers
source 1..20623
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Lam42/APPA plasmid"
BASE COUNT 5445 a 4847 c 4902 g 5429 t 0 others
ORIGIN

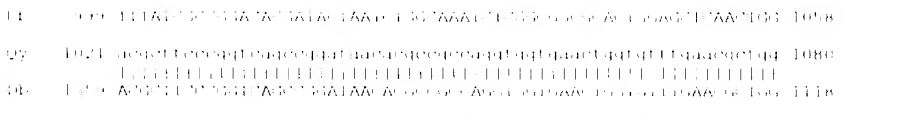
Query Match 96.9%; Score 1282.2; DB 5; Length 20623;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

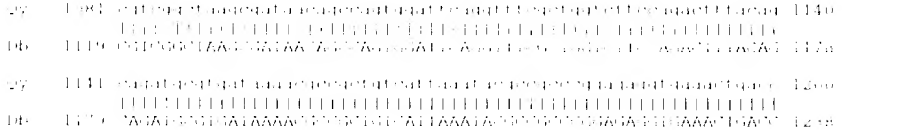


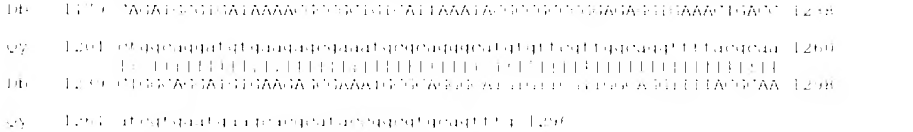
RESULT 12
AX042377
LOCUS AX042377 5421 bp DNA PAT 23-NOV-2000
DEFINITION Sequence 6 from Patent WO0064242.
ACCESSION AX042377
VERSION AX042377.1 GI:11340995
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 5421)
AUTHORS Forsberg,C.W., Solomon,S. and Phillips,J.P.
TITLE Transgenic animals expressing salivary proteins
JOURNAL Patent: WO 0064242 A 26-OCT-2000;
UNIVERSITY OF GUELPH (CA)
FEATURES Location/Qualifiers
source 1..5421
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="SV40/AIPA - Intron plasmid"
BASE COUNT 1410 a 1321 c 1442 g 1253 t 1 others

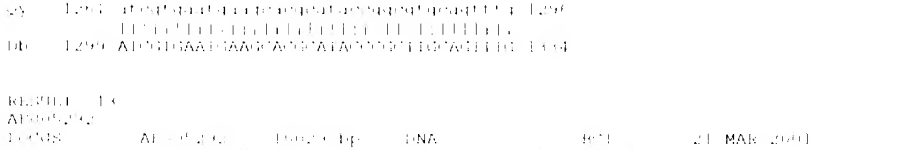
ORIGIN

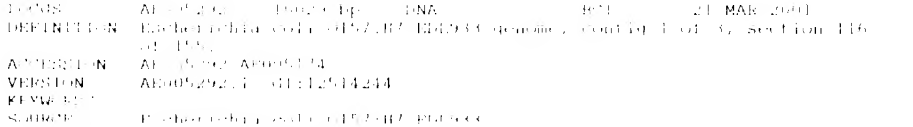
Query Match 96.0%; Score 1270.2; DB 9; Length 54...
Best Local Similarity 99.3%; Pred. No. 0;
Matches 12873; Conservative 0; Mismatches 85; Indels 1; Gaps 1;



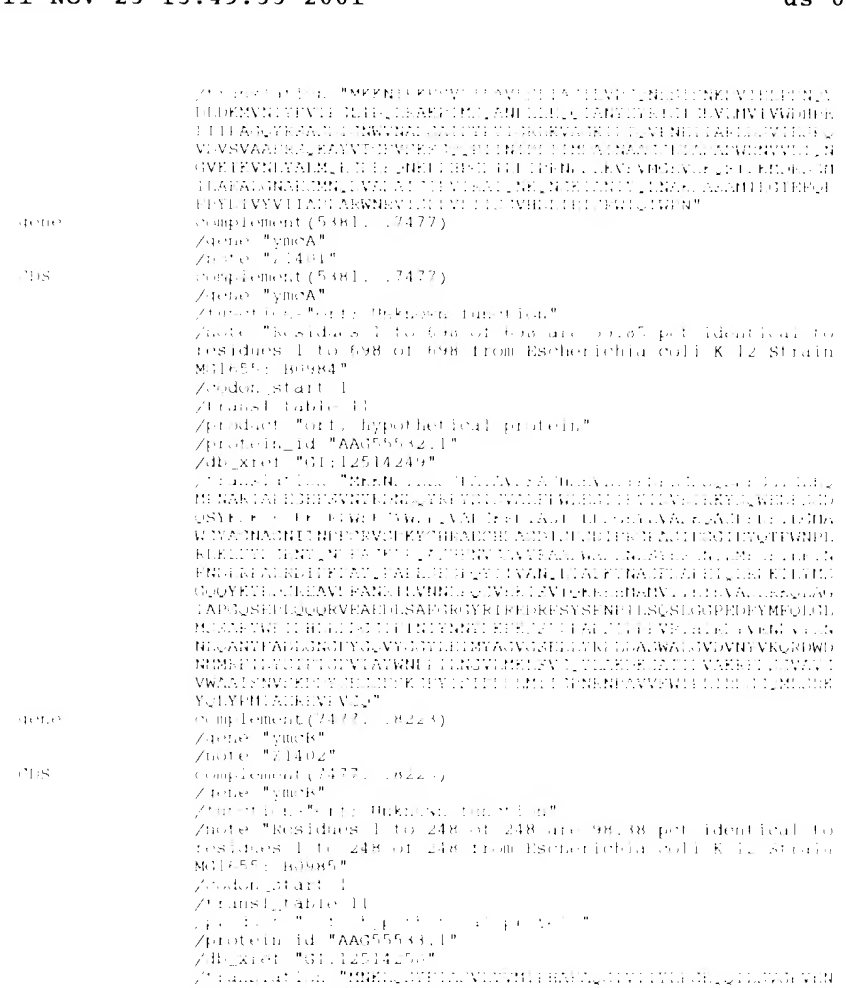




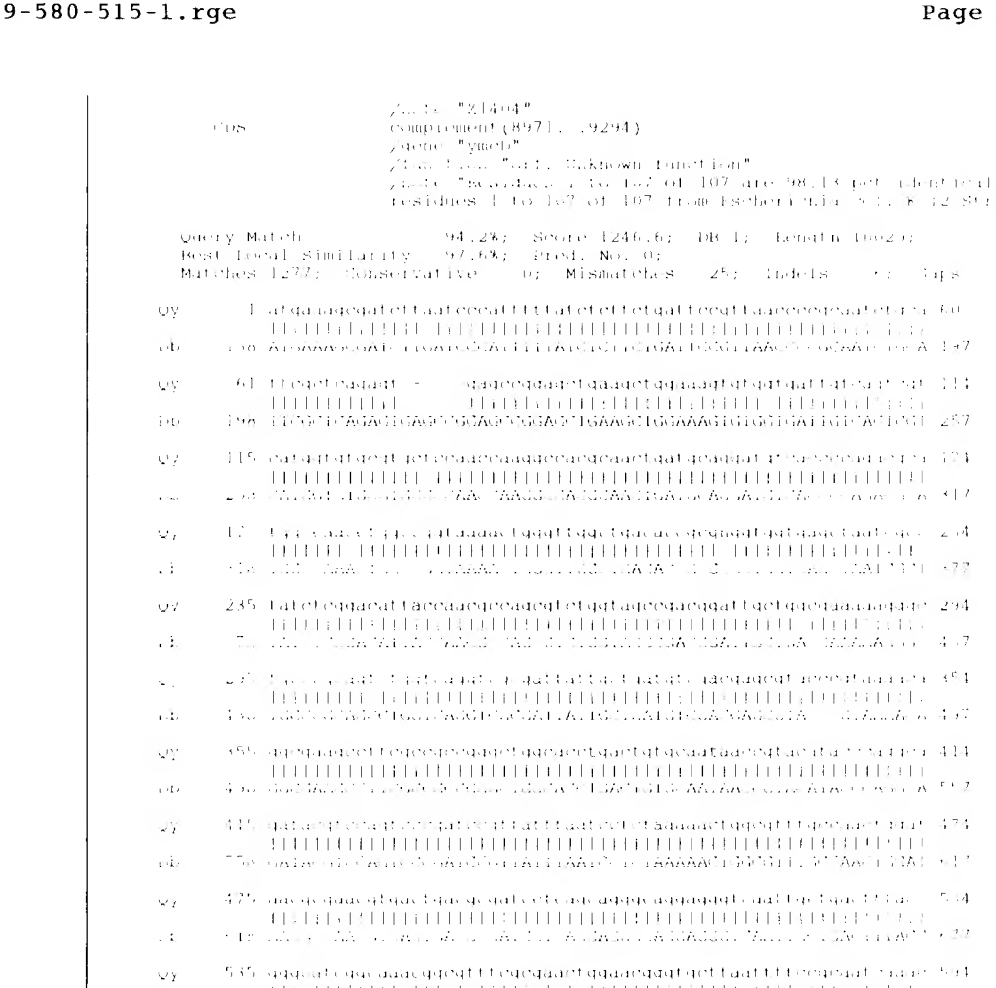




```
LOCUS       AE005292               1002 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 4, section 116
            of 155.
ACCESSION   AE005292 AP000134
VERSION     AE005292.1  GI:12514244
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; Gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1002)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E.T., Potamousis,K.,
            Apodaca,J., Anantharaman,S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21118991
    PUBMED  11206551
REFERENCE   2  (bases 1 to 1002)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E.T., Potamousis,K.,
            Apodaca,J., Anantharaman,S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1..1002
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            108..1442
                     /gene="ppA"
                     /gene_xref="GI:1742"
     CDS             108..1442
                     /gene="ppA"
                     /function="enzyme; Central intermediary metabolism; Pool,
                     multipurpose conversions of intermed. met'm"
                     /note="Residues 11 to 444 of 444 are 98.1% pet identical
                     to residues 1 to 442 of 442 from Escherichia coli K 12
                     strain MG1655; B0980"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoanhydride phosphor,ase: pt . ' ' ' ed
                     phosphatases; periplasmic"
                     /protein_id="AAG55528.1"
                     /db_xref="GI:12514245"
                     /translation="MCCMEQSNCSHEAELELLLCLGLSGVAAQSELLLKELS
                     VVEVSRSDVKAGLKALGLMGVVELAWFNWEVKLGWLLPKSGLEAVFVDVGLSVA
                     DGLLKKWQTQLSQVAIAGVGKPKLRDAFAAGLALSALVLMGLAGSNLGELEN
                     PLKIGVCQLDNANVIDALLSRAGGSIADFGHRQLAFRELPRVLNELQSNLQLNKPRQ
                     DLSQSLQAIESELEVCANNVSLCAVLAGMSLGGLLGAQMPLDWSELLQGHL
                     WNLLLSLPHNAQFYLLQRLPEVARSRALPLLLGLGMIALLFHHPQVKQAYSVLCLSVLLI
                     AGHDLNEANLSGALLLNWFLPGRPDNLPPSSDLVFLPWRKELSLNGAVQLVSLVFQLLL
                     QHRPGSLSLNFFSGVKLLALSLGLWNAQLPLSLQLVNLAGLPLA"
  gene            complement(1562..3472)
                  /gene="yeeS"
                  /note="Z1498"
  CDS             complement(1562..3472)
                  /gene="yeeS"
                  /function="orf; Unknown function"
                  /note="Residues 1 to 726 of 726 are 99.8% pet identical to
                  residues 1 to 726 of 726 from Escherichia coli K 12 strain
                  MG1655; B0981"
                  /codon_start=1
                  /transl_table=11
                  /product="orf, hypothetical protein"
                  /protein_id="AAG55529.1"
                  /db_xref="GI:12514246"
                  /translation="MSIKRNNIFPVSLFNFCLLFLVVPLAGSSSVNAFLL
                  AVAYSLLSIPIYQAPILVVVEQKLGNATLSVSLSLMIFNSSPESALFLLLQPMILGK
                  IGAELSLLDMVLQKYLPIVSIGWAILLKIKKSGELACCWMHIPSLNGQLAQDLILVGEN
                  GHYLLECSSFVNGMVSQRLSPEGVALLIAPLEAKPSLPVFVSLSLSEAINALGFIF
                  IVSIERSKESGMLELLMTGDDFQLLIRTLNGTANNYLAANIAKGAAIGSLSIFFLQMGL
                  PEVKGELDQAEEKLNVYKQAGKSVELNLLAKAVLEQTVNVSNLNLLFRLAEISQLY
                  FELGRPIYRALLPKPSLLLPSGFLLPSGFRLNKEVSAMPSTQQIVEFLPSLVLSVAVSLAPM
                  VSLSSKESSAIGNVEICIGPAVLGLFVSLAKALNVVLCLSLGSVGAVIAKAMPF
                  WSVPAPFSFTPICIVYATLFMSFWLIPPTLFLPFFNLFSNQLSPRIFLLAVLNPA
                  DSAVEAVKALFTSLHLAMMLLVNNTLMLDAIPQSPKLIVSSLAAVIAQGLKVLFI
                  SVSLDSYSHNCLVSNEGCGSLSAGGLLGNKVTQHLKLRLVLVPGVIPNHSFL
                  LMEDRMPSLLLWANCHYSVAIVLTFPMLAVSDAAVVGVSLGLLVAPFSLNLAKEVS
                  LSMSPLLSACVNLSCATLNCVLKRASLAYSVSYNYSVSSTRL"
  gene            complement(3762..4220)
                  /gene="yeeY"
                  /note="Z1499"
  CDS             complement(3762..4220)
                  /gene="yeeY"
                  /function="putative enzyme; Not classified"
                  /note="Residues 1 to 152 of 152 are 100.0% pet identical
                  to residues 1 to 152 of 152 from Escherichia coli K 12
                  strain MG1655; B0982"
                  /codon_start=1
                  /transl_table=11
                  /product="putative phosphatase"
                  /protein_id="AAG55530.1"
                  /db_xref="GI:12514247"
                  /translation="MSCNRSLSEENLLSVAVSLSGSLSLLLLLSSSSSV
                  VSSLVSHFACAIAAPSAANPVSSSSSAVSSSSSAMSNLLLLSLSLLSLSSSSSLA
                  PEVRCKIHLDPQWLDQKICIDGFYEKQLDASTEHVSIHLLPSSQLSVAKPLSK"
  gene            complement(4196..5335)
                  /gene="yeeZ"
                  /note="Z1400"
  CDS             complement(4196..5335)
                  /gene="yeeZ"
                  /function="phenotype; Not classified"
                  /note="Residues 1 to 379 of 379 are 100.0% pet identical
                  to residues 1 to 379 of 379 from Escherichia coli K 12
                  strain MG1655; B0983"
                  /codon_start=1
                  /transl_table=11
                  /product="putative function in exopolysaccharide
                  production"
                  /protein_id="AAG55531.1"
                  /db_xref="GI:12514248"
```

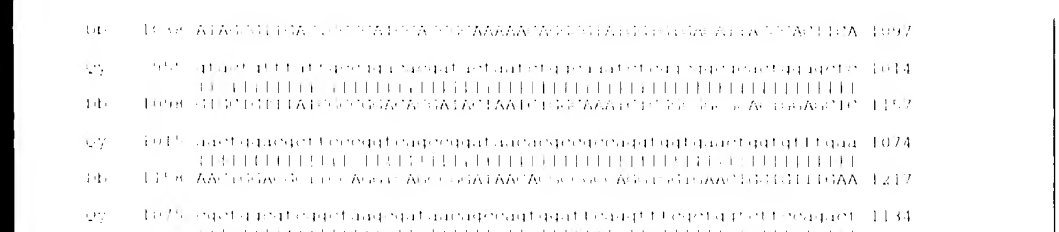


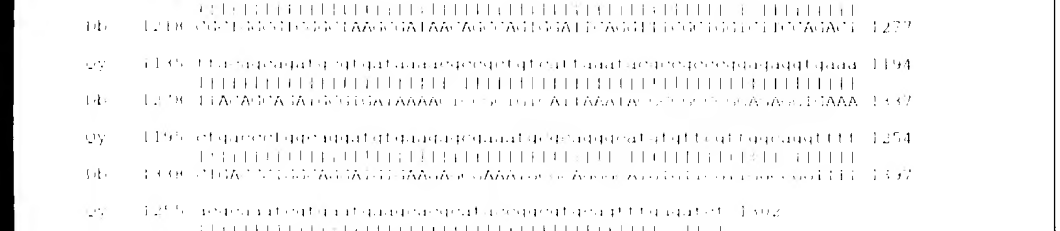
/translation="MKPNIKKVVVLGAVGVGKTAFIQRFVTGEFEENQEATIGVEFGSKIVEIDGKRIKLQIWDTAGQERFRSITRSYYRGAVGALLVYDITKRETFNHLTSWLEDARQHANSNMVIMLIGNKSDLRHLRAVPTDEARAFAEKEGLSFIETSALDSTNVEEAFKNILTEIYRIVSQKQIADRDDPAGNSSGIKVGKPVQQPTQGQAAGGCCSN"
gene complement(5381..7477)
/gene="ymcA"
/note="741d1"
CDS complement(5381..7477)
/gene="ymcA"
/function="unknown"
/product="hypothetical protein"
/note="Residues 1 to 698 of 698 are 99.85% pct identical to residues 1 to 698 of 698 from Escherichia coli K-12 Strain MG1655: b0984"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55542.1"
/db_xref="GI:12514249"
/translation="MRLNLGLIGLGRMGANMVRRLMKGGHECVVYDVNQEAVDALVKEGATGAKSLEDLVSKLEAPRVIWLMVPAGEITESVIKDLYPLLSPGDIVIDGGNSEYRDTTRRCRDLKAKGILFVGSGVSGGEEGARYGPSLMPGGQEEAWPHIKTIFQSIAAKVDGEPCCEWVGDEGAGHFVKMVHNGIEYGDMQLICEAYDVMKRVGGFSNEEMAQVFTEWNNGELDSFLIEITADIFTKKDEDGEYLVDVILDEAANKGTGKWTSQSALDLGEPLSLITESVFARCLSSLKDERVAASKVLSGPQAQPAGDKAEFIEKVRRALYLGKIVSYAQGFSQLRAASEEYNWDLNYGEIAKIFRAGCIIRAQFLQKITDAYAENPQIANLLLAPYFKQIADDYQQALRDVVAYAVQNGIPVPTFSAAVAYYDSYRAAVLPANLIQAQRDYFGAHTYKRIDKEGVFHTEWLD"
gene complement(7477..8223)
/gene="ymcB"
/note="21402"
CDS complement(7477..8223)
/gene="ymcB"
/function="orf, hypothetical protein"
/note="Residues 1 to 248 of 248 are 98.38% pct identical to residues 1 to 248 of 248 from Escherichia coli K-12 Strain MG1655: b0985"
/codon_start=1
/transl_table=11
/protein_id="AAG55543.1"
/db_xref="GI:12514250"
/translation="MDRLDFRIAVLDYDHTLAHSAQIKVNAFGHTLVELRGLEPRDFGKEFEDLFHHPFRGIEKVTGNLNNFLLREAHHGKLVCFADYVRDSIDLLDTISPSFQAVYILSNKEGFFFRDEVEARFGVGNLKDYWSRAFGCETIYHIDKLKTQGTEIIEKENPRSFLFFDIEDRELEWVNAGCAHPHCPYFFFRHGVQEDIAEALWSLFNQDAMPVLRDFREGRRPEDYVDDKAAQIFDKYLRGVEYLVSMGVDLNIPLDVQLRCIDEERFGLFGKLGPQIGECIQDVYEGRNASGVKAWVEKWMNTDVTLSRIAVHTMLLGRQENVEIPVTIIYSNETGDSQPGANELMKRRLCPVFTTRIAPLDGVMHLTRTADVVLFTLSRWVEYAAGDISLDGEWLAERDHLAQNCPLLEKIRPYLDPIAQPNQTVIDWSGSRLTAFCSTKGNRFNAISNMLDGGETLVTCFGVDVPMDIKLRGHHLTGKDSGESAKGFKIVVRNIFPVYRSEDYNYDLYFTVQRFVITLLGAVSGAGQLPWGAIKGVTIDLIRTAGALENNVLTLTEKVVAERVALWNKSHVEELPPGSGLWIGFSAHVLPEKYADLNDQIVSVLTQRVPD"
gene complement(8220..8864)
/gene="ymcC"
/note="21403"
CDS complement(8220..8864)
/gene="ymcC"
/function="putative regulator; Not classified"
/note="Residues 1 to 214 of 214 are 99.5% pct identical to residues 1 to 214 of 214 from Escherichia coli K-12 Strain MG1655: b0986"
/codon_start=1
/transl_table=11
/product="putative regulator"
/protein_id="AAG55544.1"
/db_xref="GI:12514251"
/translation="MRTLSAIIAAGVTAGLSAWGLTAAIFNGAIVAVGLEFLGRPDQLSPVFVSGSIIRMGAVTVVGCLFLSIPVFRLNNELRVLNGNLLSGLGLAFAFGGWFDLFSLHIGAWTAFLGSLLYVVYIGFSVSLLLQLNVWGEVLTPFKMGNTFWVFAALVPGCFLPLMDTVFVSGTFRAP"
gene complement(8971..9294)
/gene="ymcD"

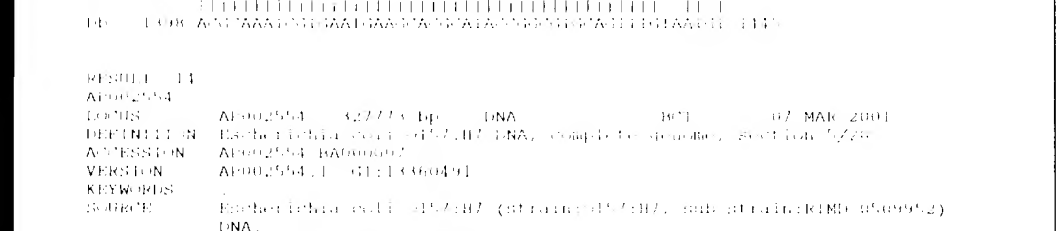


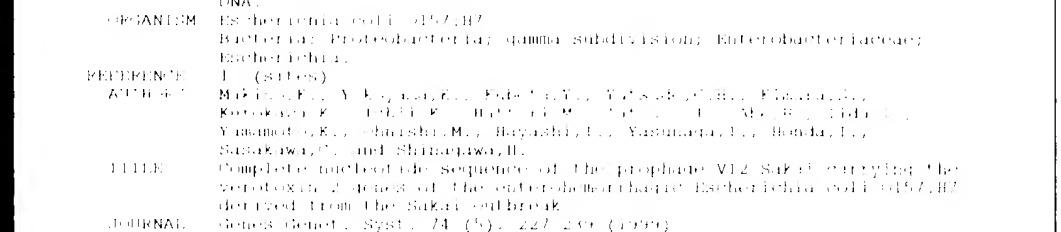
/translation="MIGLTGGIGSGKSTVANLFTELGVPVIDADIIARQVVEPGQPALAEIVRHFGTDILQADGSLNRAALRERIFANPEQKKQLEAILHPLIAEETAWQAAQAESPYCILVIPQLLETGSYRSLVDRILVVDVPEQLQIERTMKRDGSTEAQVRAIMATQVSREQRLAVADDVIDNSGTLEELEQKVDKLDKKYLELAANKNL"
gene complement(8971..9294)
/gene="ymcD"
/note="21404"
CDS complement(8971..9294)
/gene="ymcD"
/function="orf, hypothetical protein"
/note="Residues 1 to 107 of 107 are 98.13% pct identical to residues 1 to 107 of 107 from Escherichia coli K-12 Strain MG1655: b0987"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55545.1"
/db_xref="GI:12514252"
/translation="MIGLTGGIGSGKSTVANLFTELGVPVIDADIIARQVVEPGQPALAEIVRHFGTDILQADGSLNRAALRERIFANPEQKKQLEAILHPLIAEETAWQAAQAESPYCILVIPQLLETGSYRSLVDRILVVDVPEQLQIERTMKRDGSTEAQVRAIMATQVSREQRLAVADDVIDNSGTLEELEQKVDKLDKKYLELAANKNL"

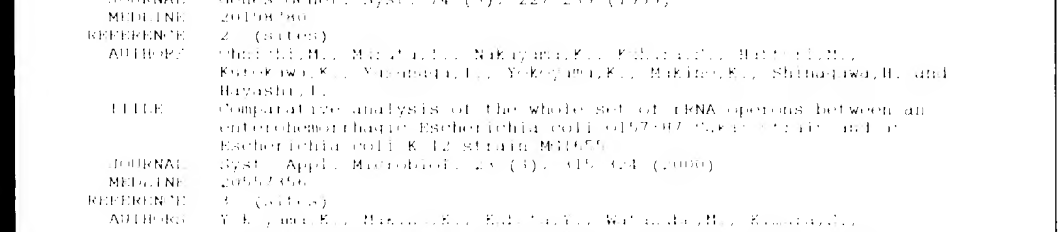
Query Match 94.2%; Score 1246.6; DB 1; Length 10020;
Best Local Similarity 97.6%; Prod. No. 0;
Matches 1277; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

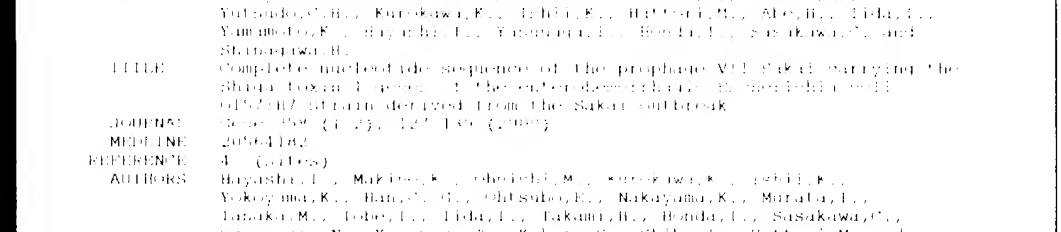






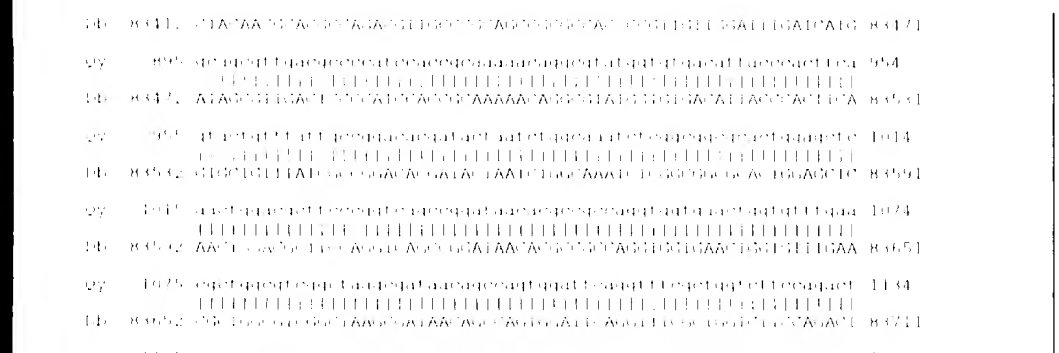






RESULT 14
AP002564
LOCUS AP002564 42777 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 7/20.
ACCESSION AP002564 BA000007
VERSION AP002564.1 GI:13360491
KEYWORDS .
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)
AUTHORS Makino,K., Yamamoto,K., Fukushi,Y., Yokoyama,K., Nakayama,Y.,
Kurokawa,K., Ishii,K., Hattori,M., Tamura,K., Tobe,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
JOURNAL Genes Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE 20190790
REFERENCE 2 (sites)
AUTHORS Ohnishi,M., Makino,K., Nakayama,K., Fukushi,Y., Hattori,M.,
Kurokawa,K., Yamamoto,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
TITLE Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 [corrected] and a
Escherichia coli K-12 strain MG1655
JOURNAL Syst. Appl. Microbiol. 24 (3), 415-424 (2001)
MEDLINE 20557356
REFERENCE 3 (sites)
AUTHORS Yamamoto,K., Makino,K., Fukushi,Y., Watanabe,H., Kimata,K.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
JOURNAL Gene 258 (1-2), 127-139 (2000)
MEDLINE 20564182
REFERENCE 4 (sites)
AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL DNA Res. 8 (1), 11-22 (2001)
MEDLINE 21156241
REFERENCE 5 (bases 1 to 42777)
AUTHORS Ohnishi,M., Kurokawa,K., Makino,K., Yamamoto,K., Yutsudo,C.H. and
Hayashi,T.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen.info.osaka-u.ac.jp,
URL:http://www.gen.info.osaka-u.ac.jp/, Tel:+81-6-6879-4067,
Fax:+81-6-6879-4047)
COMMENT genome project.
FEATURES Location/Qualifiers
source 1..42777
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
gene 150..2258
/gene="ECs1032"
CDS 150..2258
/gene="ECs1032"
/note="similar to YCRY ECOLI|P37181 hypothetical protein in
99 in 702 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative oxidoreductase"
/protein_id="BAB34455.1"
/db_xref="GI:13360492"
/translation="MNSLFASIASCLEELFKIEENESAVEVLVVGVPFEKLIKLV
YQSLMWSRLASRIMLPLGHCKVYSDLPLYLGVQAINWIEMPNESATFAVHLGGLNEII
RNQQYGAMEVEDAIVDAFTEKNLPEPNVEDAIS[AVNVWCRFTLNSIAEELSWRFLH
LRGYRDRAGIAPIKELLAAAIVMRSGWLPGTLPLLDPMGGSGILTLEAAMLAIDRAPGL
HRGRWGPSGWAQHDEAIWQEVKAEAQTRARKGLALVSHHTGSGLGAKVRVLGRAKNAP
LAGIGELIIFEVKDVAQLINPLPKEPYGIVLSNEPYSEPELSEEAITALHSLLGKIMF
NQPSGWNLGLPSASPELLSPLQLRADKQYKAKNCELPVQKNVHVALSLEGKSAMVA
IEVANwwRKNLKKEIRWAKQECIEGTRLYDAPLLEYNVAVGVGAVAVVVLVAPPKEII
PGRPAEQSIGLAAGLVEQLALRPVLPNPEPMQLLPGLAELSEEAITALHSLLGKIMF
HLWVNLTDYLDIGLFLDHRIARRMLQQMSKGKDFLNLESVIGSAIVHAGLGGAKSLII
VDMSRTYLDWAERNLRLNGLPGRAHKLLQAGCLAWLRARPELPLLTPSPGKKR
MEDAPDVPPHLALMKELPELLPAGSTIMFSNNEPGGFEMDLLPGLAELSEEAITALHS
LSQPFAKNRQTHNGWLTIAA"
gene 2270..4177
/gene="ECs1033"
CDS 2270..4177
/gene="ECs1033"
/note="similar to HUP ECOLI|P17182 hypothetical protein in
635 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative ATP-binding component of a transport
system"
/protein_id="BAB34456.1"
/db_xref="GI:13360493"
/translation="MSCCLRHEGAKLAFGGAVILNGADFTLNEGVVTVVVGPNGAGKST
LMRLLNGLVLPSEGTIYIGQKLIVAEDQGLDPEFPNVFRCVVDEVAFGPVNLKVP
KERKQVMNLVKEMLNGNLILAEVLPQLLIHEGNWKLSESLNRVAAFFTLMDIVVVLGK
WLEPKAAIGKALVSNFKVLLLDEPTNHLDIETIEWLEGFLKTFKGTVIVISHDRWFLN
NATEWVDLDRGKLVYYPNNYQQYLEEKEERLAEEQKQNAELEKFIERFRYKATKAKQ
KSRIKMLEKMELIEVVHVDSPFAFTFRFENQRPGKVIMELNHVSFAYPDQPVLFKDVN
VKDFSAQVLRGDKTALIGPNGCGKTTLLKLMLGQLQADSGRIHVGTKLEVAYFDQHRA
ELLFEETVMDNLAEDAQEAVMVNGRPRHVLGYLQEFLFSPKRAMTPVKALSGGERNRL
LARLFLKPSNLLILDEPTNDLDVETLELLEELIDSYQGTVLLVSHDRQFVDNIVTEIW
TFEGDGKIGRYVGGIHDARGQQEQEVVALEQPAVEKNEEPAAPEKAEEVEKPQSYKKLSYKL
QRELEQLPQLIEDLESAKLEAELQTQVADPASFTSQTHEQLQFVLADMAAAEQELEQAEEA
WEYLELALKNGC"
gene 4307..5560
/gene="ECs1034"

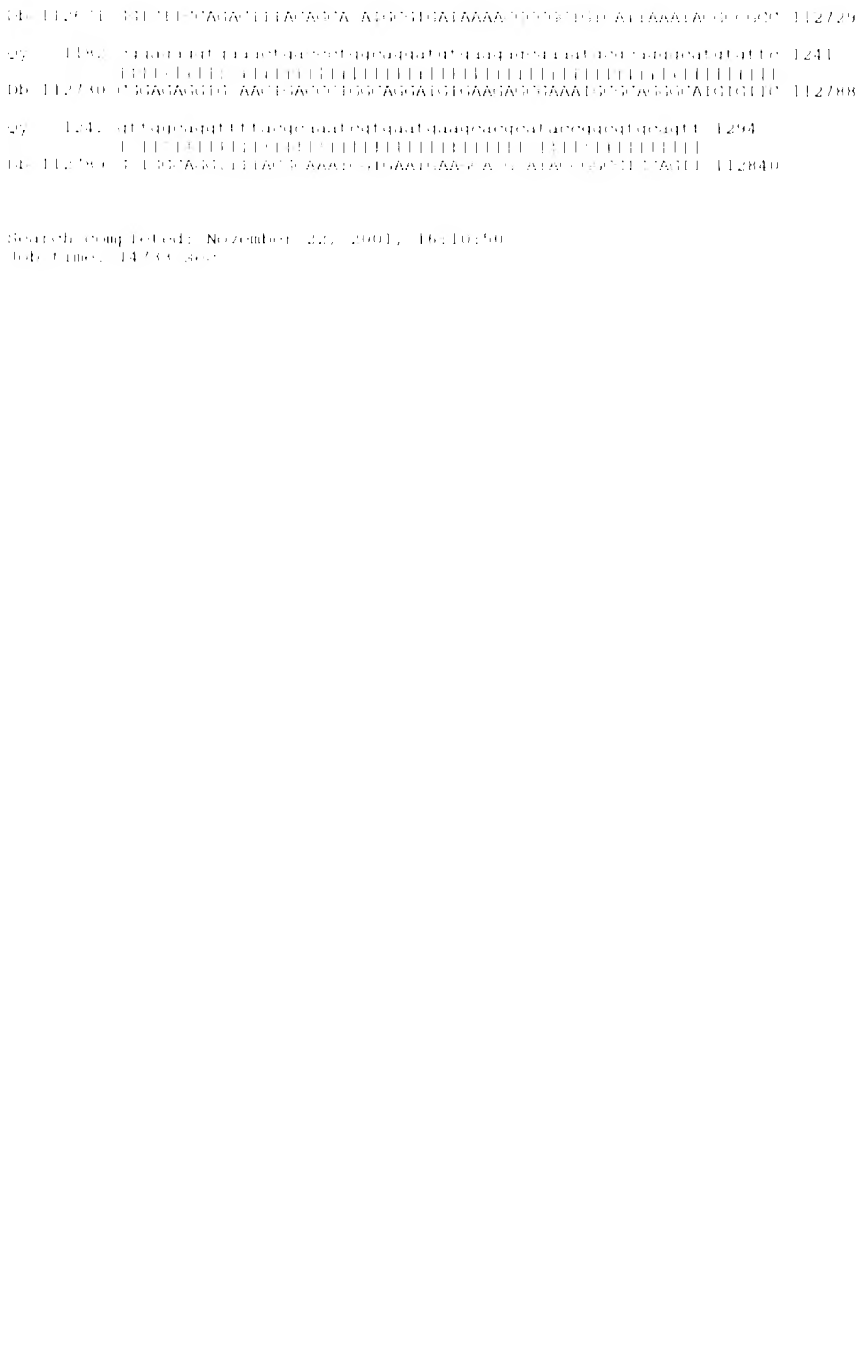
```
     CDS             4...7...8566...
                     /gene="EG10434"
                     /note="similar to PqiA_ECOLI:q11787185 percent identity
                     97 in 477 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="paraquat inducible protein A"
                     /protein_id="BAB34467.1"
                     /db_xref="GI:13360042"
                     /translation="MVTHEHAAEKIVVLVGIVLAVLIVGAVLFSLKAASISPVAW
                     LAPNQGLAYAVAMAFMGLFSNLLRSVNMNVWMVSLSLFLLGIGAVLESLVASSISL
                     ELLEVGLVPAFCLLIILLLVNEALLRVRLRDLLAPVLILLKIGMGAICFLAGVLVGV
                     KLDSAVLVVVVLVLGLGLSFVLGPLVTRLLLGYEYDPEGEVSPGTLAKTGIQGGLK
                     GLLKSCSCQIATLLPALPLPVCFRCVERKRNCLAWTLALLVLGLMLVLLANTLFM
                     VIDLDSKPMVLLLAVLLLRRPLGTVFVSAVLFLASCMVFLKMLAGMGVWSAGHC
                     KCPLFPMGLLVFVVIVVVDWGMGVLVVLSVLSVSMSLMLVFLAEKCHLADVVCM
                     IPLSAMLVDLDLSWWSQGFLVLLE"
     gene            7...7205
                     /gene="EG10435"
     CDS             7...7205
                     /gene="EG10435"
                     /note="similar to PqiB_ECOLI:q11782184 percent identity
                     99 in 546 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="paraquat inducible protein B"
                     /protein_id="BAB34468.1"
                     /db_xref="GI:13360430"
                     /translation="MFSNRIAPLKVFWLCVAVLLVVALLSAWVLFTHYSHGPE
                     VIIIDAGWAGDKSLKSVAVVALRAGLPVGEVKSIDLDAHLI
                     VEWVVKPLYKDSISKDSKMKKYLVSSPDLADMSAKLIRV
                     VELQPPSQLTGQTLVFLDPVVKGVPPTFIEPRSLGINAETGELLGVSL
                     WKEPLAVVSLALPPVLMSGLEGVPPAFIEAPELAMPVLVRVLSS
                     LQDSLYLIHHLYLMLPKPVPGLPLSAVLPSGVPLVSAVDPNGQTA
                     IPVLLKLSPELRMKLGDNAFVVPPLPLKPFLPSVLPLTGNLVTLALYVDLDFTPNI
                     PATGISNAGNQGTLVGGLLVSMPFLLPLGLGLLWNLLVRLMQRL
                     MFRQLIIIAGGELIASLPMLLVLFLMLGLNFGLQLPLGGLASVAAGMQKL
                     DQVLRFLQPVLRLLNPKSNALVFPAKPFLDPPRPAQ"
     gene            7...7765
                     /gene="EG11036"
     CDS             7217...7765
                     /gene="EG11036"
                     /note="similar to YBBA_ECOLI:q11787185 percent identity
                     98 in 182 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB34469.1"
                     /db_xref="GI:13360496"
                     /translation="MSIAGALLAGLLGMFGLIVVGLGAAGHELAVE
                     QVAVLSLAIPLASSAQDFLNVIVIFLNGIGSAGFVLGHAFLSGL
                     EVLAAVWGLFAACIALFRRLP"
     gene            8021...8188
                     /gene="EG11037"
     CDS             8021...8188
                     /gene="EG11037"
                     /note="identical to BBB_ECOLI:q11787186 (Conserved in
                     E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="thioredoxin-like protein"
                     /protein_id="BAB34470.1"
                     /db_xref="GI:13360497"
                     /translation="MFRQLIIIAGGELIASLPMLLVLFLMLGLNFGLQLPLGGLASVAAGMQKL
                     EEZMCDVVMA"
     gene            complement(8258...8776)
                     /gene="EG11038"
     CDS             complement(8258...8776)
                     /gene="EG11038"
                     /note="similar to PABA_ECOLI:q11787187 protein of unknown
                     100 in 172 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="beta-hydroxydecanoyl-thioester dehydrase"
                     /protein_id="BAB34471.1"
                     /db_xref="GI:13360498"
                     /translation="MVERRESVIEEILAAFSKVSLSDMKPDASFVDKM

  Query Match             94.2%;  Score 5246.5;  DB 2;  Length 8776;
  Best Local Similarity   97.6%;  Pred. No. 0;
  Matches  5...  Mismatches   25;  Indels    2;  Gaps    1;
```



RESULT 15
AC020970
LOCUS AC020970 21236 bp DNA HTG 10-FEB-2000
DEFINITION Mus musculus clone RP23-252M21, WORKING DRAFT SEQUENCE, 144 unordered pieces.
ACCESSION AC020970
VERSION AC020970.1 GI:6691260
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 21236)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 21236)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web [illegible]
Summary Statistics
Estimated insert size: 21236; sum of contigs estimation
Estimated insert size: 170000, pulse field gel estimation
Quality coverage: 4.29X in Q20 bases; pulse field gel estimation
Quality coverage: 3.10X in Q20 bases; sum of contigs estimation

* NOTE: This is a "working draft" sequence. It currently
* consists of 144 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1159: contig of 1159 bp in length
* gap of unknown length
* 1160 2363: contig of 1204 bp in length
* gap of unknown length
* 2364 3604: contig of 1241 bp in length
* gap of unknown length
* 3605 5188: contig of 1584 bp in length
* gap of unknown length
* 5189 6472: contig of 1284 bp in length
* gap of unknown length
* 6473 7837: contig of 1365 bp in length
* gap of unknown length
* 7838 9783: contig of 2046 bp in length
* gap of unknown length
* 9784 11046: contig of 1263 bp in length
* gap of unknown length
* 11047 12223: contig of 1177 bp in length
* gap of unknown length
* 12224 13492: contig of 1269 bp in length
* gap of unknown length
* 13493 14669: contig of 1177 bp in length
* gap of unknown length
* 14670 15886: contig of 1217 bp in length
* gap of unknown length
* 15887 17089: contig of 1203 bp in length
* gap of unknown length
* 17090 18308: contig of 1219 bp in length
* gap of unknown length
* 18309 19518: contig of 1210 bp in length
* gap of unknown length
* 19519 20710: contig of 1192 bp in length
* gap of unknown length
* 20711 21932: contig of 1222 bp in length
* gap of unknown length
* 21933 23606: contig of 1674 bp in length
* gap of unknown length
* 23607 25328: contig of 1722 bp in length
* gap of unknown length
* 25329 26649: contig of 1321 bp in length
* gap of unknown length
* 26650 27849: contig of 1200 bp in length
* gap of unknown length
* 27850 29135: contig of 1286 bp in length
* gap of unknown length
* 29136 30700: contig of 1565 bp in length
* gap of unknown length
* 30701 31877: contig of 1177 bp in length
* gap of unknown length
* 31878 33065: contig of 1188 bp in length
* gap of unknown length
* 33066 34267: contig of 1202 bp in length
* gap of unknown length
* 34268 35474: contig of 1207 bp in length
* gap of unknown length
* 35475 36650: contig of 1176 bp in length
* gap of unknown length
* 36651 37859: contig of 1209 bp in length
* gap of unknown length
* 37860 39072: contig of 1213 bp in length
* gap of unknown length
* 39073 40406: contig of 1334 bp in length
* gap of unknown length
* 40407 41630: contig of 1224 bp in length
* gap of unknown length
* 41631 42864: contig of 1234 bp in length
* gap of unknown length
* 42865 44200: contig of 1336 bp in length
* gap of unknown length
* 44201 46087: contig of 1887 bp in length
* gap of unknown length
* 46088 48227: contig of 2140 bp in length
* gap of unknown length
* 48228 49518: contig of 1291 bp in length
* gap of unknown length
* 49519 50910: contig of 1392 bp in length
* gap of unknown length

```
*     [illegible]   52426: contig of 1510 bp in length
*                          gap of unknown length
*     52427   53615: contig of 1189 bp in length
*                          gap of unknown length
*     53614   55004: contig of 1391 bp in length
*                          gap of unknown length
*     55005   56300: contig of 1296 bp in length
*                          gap of unknown length
*     56301   57505: contig of 1205 bp in length
*                          gap of unknown length
*     57506   58686: contig of 1181 bp in length
*                          gap of unknown length
*     58687   60407: contig of 1721 bp in length
*                          gap of unknown length
*     60408   61682: contig of 1275 bp in length
*                          gap of unknown length
*     61683   62881: contig of 1199 bp in length
*                          gap of unknown length
*     62882   64220: contig of 1340 bp in length
*                          gap of unknown length
*     64221   65420: contig of 1200 bp in length
*                          gap of unknown length
*     65421   66605: contig of 1185 bp in length
*                          gap of unknown length
*     66606   68682: contig of 2077 bp in length
*                          gap of unknown length
*     68683   70712: contig of 2030 bp in length
*                          gap of unknown length
*     70713   71895: contig of 1183 bp in length
*                          gap of unknown length
*     71896   73462: contig of 1567 bp in length
*                          gap of unknown length
*     73463   74806: contig of 1344 bp in length
*                          gap of unknown length
*     74807   76205: contig of 1399 bp in length
*                          gap of unknown length
*     76206   77580: contig of 1375 bp in length
*                          gap of unknown length
*     77581   79081: contig of 1501 bp in length
*                          gap of unknown length
*     79082   80314: contig of 1233 bp in length
*                          gap of unknown length
*     80315   81899: contig of 1585 bp in length
*                          gap of unknown length
*     81900   83140: contig of 1241 bp in length
*                          gap of unknown length
*     83141   84665: contig of 1525 bp in length
*                          gap of unknown length
*     84666   85888: contig of 1223 bp in length
*                          gap of unknown length
*     85889   87155: contig of 1267 bp in length
*                          gap of unknown length
*     87156   88435: contig of 1280 bp in length
*                          gap of unknown length
*     88436   89955: contig of 1520 bp in length
*                          gap of unknown length
*     89956   91277: contig of 1322 bp in length
*                          gap of unknown length
*     91278   92513: contig of 1236 bp in length
*                          gap of unknown length
*     92514   93917: contig of 1404 bp in length
*                          gap of unknown length
*     93918   95448: contig of 1531 bp in length
*                          gap of unknown length
*     95449   96800: contig of 1352 bp in length
*                          gap of unknown length
*     96801   97978: contig of 1178 bp in length
*                          gap of unknown length
*     97979   99178: contig of 1200 bp in length
*                          gap of unknown length
*     99179   [illegible]: contig of 1248 bp in length
*                          gap of unknown length
*    100418   101827: contig of 1390 bp in length
*                          gap of unknown length
*    101928   103117: contig of 1190 bp in length
*                          gap of unknown length
*    103118   104550: contig of 1433 bp in length
*                          gap of unknown length
*    104551   105854: contig of 1304 bp in length
*                          gap of unknown length
*    105855   107524: contig of 1669 bp in length
*                          gap of unknown length
*    107525   109208: contig of 1685 bp in length
*                          gap of unknown length
*    109209   110430: contig of 1222 bp in length
*                          gap of unknown length
*    110431   111662: contig of 1232 bp in length

 Query Match              [illegible]%; Score 713.4; DB 6;  Length [illegible];
 Best Local Similarity    97.3%;  Pred. No. 3.7e-195;
 Matches  922;  Conservative    0;  Mismatches    6;  Indels    [illegible];  Gaps   [illegible];

Qy      842 tacccgtaaaacaqqqcqaaqcttcqcccqcccqqactqqcacctqactqtqccaataacqqt 901
            ||||||||||| ||| ||||||||||||||| |||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy      902 qcqatacccaaccaaatacqtcaatcccqatcqttatttaatcctctaaacctcaaaqtcqcc 961
            ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy      962 ttqccaactqqatcaccqqaacqtqactqacqccqatcctcqccqaatqccqaaatqcttttt 1021
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1022 tactqacttttacccqqcatccqccaaccqqttccqccqaactqaaaccqqatqcttttttt 1081
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1082 tqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttt 1141
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1142 qatcttccaqqcttcaccqqcaatqctqctqataaacqccqctatcattaacqqtatcqtt 1201
            |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1202 qatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttttqccaac 1261
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1262 tqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttttqcc 1321
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1322 aactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttttq 1381
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1382 ccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttt 1441
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1442 tqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqctt 1501
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1502 ttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqc 1561
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1562 ttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatq 1621
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1622 cttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqat 1681
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1682 qcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqq 1741
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1742 atqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaacc 1801
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1802 qqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqaaa 1861
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1862 ccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqactqa 1921
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1922 aaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqact 1981
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     1982 qaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqqa 2041
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2042 ctqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqccqq 2101
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2102 actqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccqcc 2161
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2162 qqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttccq 2221
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2222 ccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttcc 2281
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2282 qccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqttc 2341
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2342 cqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqqtt 2401
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2402 ccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqcqq 2461
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2462 ttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcqqc 2521
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2522 qqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqqcq 2581
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2582 qcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqqcqq 2641
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2642 cqcqqcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqcqtqq 2701
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2702 cqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqqc 2761
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2762 qtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcqq 2821
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2822 cqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqcq 2881
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2882 qcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqqc 2941
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     2942 qqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcqq 3001
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     3002 cqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqcq 3061
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     3062 qcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaqc 3121
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]

Qy     3122 qqcqqcqtqqcqqcqqcqqttccqccqqactqaaaccqqatqcttttqccaactqqatcaq 3181
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   [illegible] [illegible] [illegible] [illegible]
```



Search Completed: November 23, 2001, 16:10:50
Job Time: 1473 secs

SW protein - protein search, using sw model

Run on: November 23, 2001, 15:50:48 ; Search time 21.25 Seconds
(without alignments)
43.549 Million cell updates/sec

Title: US-09-580-515-2
Perfect score: 1462
Sequence: 1 MKATLIPFLSLLIPLQSA.....VNEARELAVSLESHHHHHH 447

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 42526 seqs, 13295127 residues

Total number of hits satisfying chosen parameters: 42526

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Prel. No. is the number of results predicted by homology to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1502	6.5%	561	5	Q96NZ8	Q96NZ8 leishmania
2	1400	6.3	542	5	Q00BA2	Q00BA2 leishmania
3	1333	5.8	395	5	Q9VD68	Q9VD68 drosophila
4	1333	5.8	416	5	Q19390	Q19390 caenorhabdi
5	1241.5	5.4	362	5	Q9VVZ9	Q9VVZ9 drosophila
6	1180	5.2	365	8	Q9UUX2	Q9UUX2 aspergillus
7	1176.5	5.1	442	5	Q00UU3	Q00UU3 drosophila
8	1175.5	5.0	447	5	Q00UV1	Q00UV1 drosophila
9	1173.5	5.0	414	5	Q29KA2	Q29KA2 caenorhabdi
10	1172.5	5.0	442	5	Q97I68	Q97I68 drosophila
11	1172.5	5.0	447	5	Q97I87	Q97I87 drosophila
12	1172.5	5.0	447	5	Q00UU9	Q00UU9 drosophila
13	1172.5	5.0	447	5	Q9U5A7	Q9U5A7 drosophila
14	1172.5	5.0	447	5	Q00UU5	Q00UU5 drosophila
15	1172.5	5.0	447	5	Q00UU2	Q00UU2 drosophila
16	1172.5	5.0	447	5	Q00UU0	Q00UU0 drosophila
17	1172.5	5.0	447	5	Q00UU8	Q00UU8 drosophila
18	1172.5	5.0	447	5	Q00UU6	Q00UU6 drosophila
19	1172.5	5.0	447	5	Q00UU5	Q00UU5 drosophila
20	1172.5	5.0	447	5	Q00UU2	Q00UU2 drosophila
21	1172.5	5.0	447	5	Q00UU1	Q00UU1 drosophila
22	1172.5	5.0	447	5	Q9IW75	Q9IW75 drosophila
23	1172.5	5.0	447	5	Q9IW40	Q9IW40 drosophila
24	1172.5	5.0	447	5	Q9IW19	Q9IW19 drosophila
25	1172.5	5.0	447	5	Q9IW17	Q9IW17 drosophila
26	1172.5	5.0	447	5	Q9IVX2	Q9IVX2 drosophila
27	1172.5	5.0	447	5	Q9IVH9	Q9IVH9 drosophila
28	1171.5	5.0	447	5	Q00UV0	Q00UV0 drosophila
29	1171.5	5.0	447	5	Q00UU4	Q00UU4 drosophila
30	1171.5	5.0	447	5	Q00UU7	Q00UU7 drosophila
31	1171.5	5.0	447	5	Q00UU3	Q00UU3 drosophila
32	1170.5	5.0	447	5	Q00UU8	Q00UU8 drosophila
33	1170.5	5.0	447	5	Q00UU1	Q00UU1 drosophila
34	1170.5	5.0	447	5	Q00UU9	Q00UU9 drosophila
35	1170.5	5.0	447	5	Q00UU0	Q00UU0 drosophila
36	1170.5	5.0	447	5	Q00US9	Q00US9 drosophila
37	1170.5	5.0	447	5	Q00US8	Q00US8 drosophila
38	1170.5	5.0	447	5	Q9IVM1	Q9IVM1 drosophila
39	1170	4.8	684	5	O00838	O00838 leishmania
40	1170	4.8	707	5	O00839	O00839 leishmania
41	1169.5	4.8	447	5	Q97I86	Q97I86 drosophila
42	1169.5	4.8	447	5	Q00UU6	Q00UU6 drosophila
43	1167.5	4.8	447	5	Q00UT4	Q00UT4 drosophila
44	1165	4.7	2564	5	Q96VV7	Q96VV7 drosophila
45	1158.5	4.7	10223	2	Q54296	Q54296 streptomyce

ALIGNMENTS

RESULT 1
Q96NZ8
ID Q96NZ8 PRELIMINARY; PRT; 561 AA.
AC Q96NZ8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MEMBRANE BOUND ACID PHOSPHATASE 2.
GN XMBAP2.
OS Leishmania mexicana.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MNYC/BZ/62/M379;
RA Weise F., Wiese M.;
RT "Identification of a second membrane bound acid phosphatase of Leishmania mexicana.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ303099; CAC19581.1; -.
SQ SEQUENCE 561 AA, 58294 MW, 43EF7929E2B13798 CRC64;

Query Match 6.5%; Score 1502; DB 5; Length 561;
Best Local Similarity 28.2%; Pred. No. 0.0006;
Matches 106; Conservative 64; Mismatches 197; Indels 96; Gaps 10;

Qy 1 MKATLIPFLSLLIPLQSAFAQSEPELKLESVVIVSMGYVKAPLKA 46
Db 1 MPQVEIFGIIAAFLLVFALVIAAFLMKLVMVQLLRRHFAWIAPFSYNKLQLNR 56

Qy 47 AWPTWEVEIWQLTPKQGILAYLGHYIGLILVAGSLCAKEQPGLQVAIIALVD 98
Db 57 GYLIWS-GIEMLGKLSAFLRSKYNILAVVSEPMFESEDA-LVLVAVGG 104

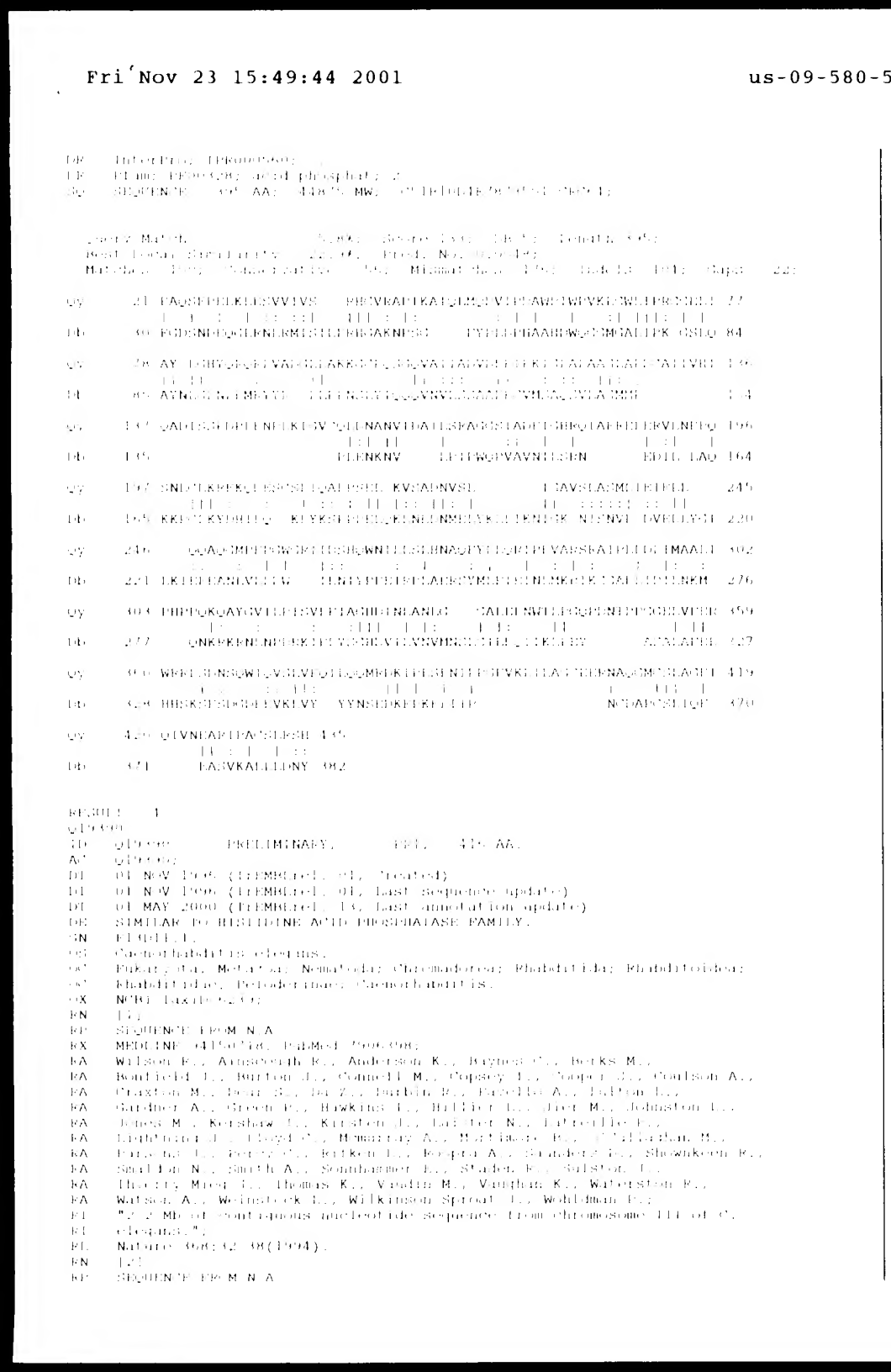
Qy 99 EKEKEIGTAFAAGLAPSTALVHLQADIGSPSPFLNPLKIGVSLLNANVHIATES 152
Db 105 SIDVQKHLQSAESFLRGFFPNL-KSEYFATHIVETQQAYTLYLNYVLQ 154

Qy 153 KAGYSAVGLMSGGVKAVPSFVVDK-NPTLFGIL-IPLAGFVLSSTGS-DLFT 204
Db 155 PQPYWGLLMAGVKAVPSFVVDKANPTLFGIL-IPLAGFVLSSTGSPDLFT 204

RESULT 2
Q94IA2
ID Q94IA2 PRELIMINARY; PRT; 542 AA.
AC Q94IA2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PROBABLE LYSOSOMAL ACID PHOSPHATASE.
GN LMJF_12_719.12.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN FRIEDLIN;
RA Peacock M., Cahel C., Mueller-Auer S., Colmenero M., Ricker M.,
RA Ivens A.C., Lawson D., Quail M., Rajandream M.A., Barrell B.G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN FRIEDLIN;
RX MEDLINE=21141495; PubMed=11077211;
RA Ivens A.C., Lewis S.M., Bontempi E., Zhang L., Chan M.M.,
RA Smith D.F.,
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:1225-1235(1998).
DR EMBL; AL157245; CAB55417.1; -.
DR InterPro; IPR000560; -.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; UNKNOWN_1.
SQ SEQUENCE 542 AA; 59618 MW; 07427F1C2E1C6C74 CRC64;

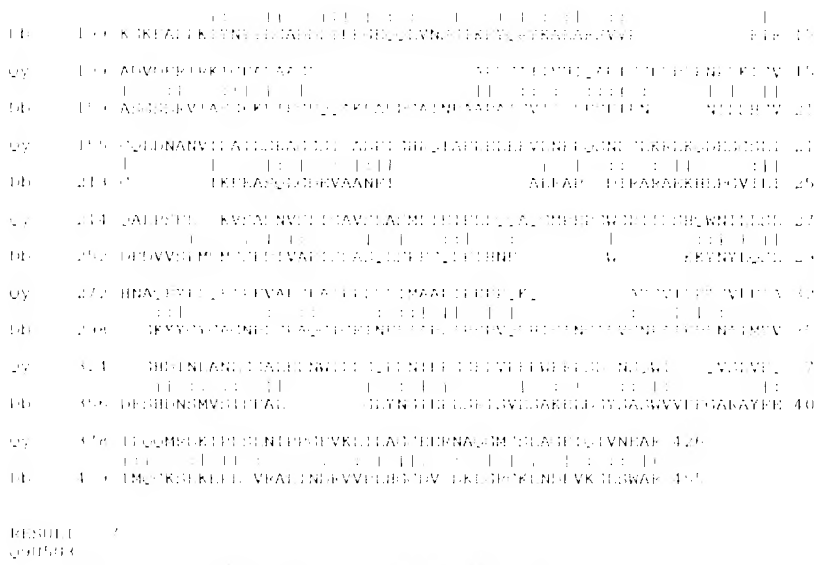
Query Match 5.1%; Score 148; DB 1; Length 542;
Best Local Similarity 22.8%; Pred. No. 6.9e-19;
Matches 152; Conservative 69; Mismatches 192; Indels 89; Gaps 19;

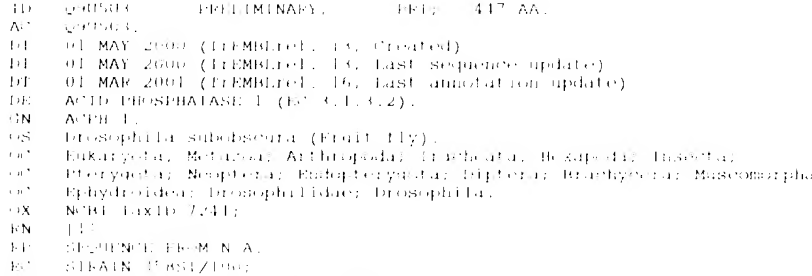
RESULT 3
Q9VD68
ID Q9VD68 PRELIMINARY; PRT; 895 AA.
AC Q9VD68;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-SEP-2000 (TrEMBLrel. 15, Last annotation update)
DE CG6656 PROTEIN.
GN CG6656.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN BERKELEY;
RX MEDLINE=20194095; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA [illegible] Weinstock G.M., Venter J.C., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C.,
RA Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Hoxtwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhou X., Zhong F.N., Zhong W., Zhu X., Zou P., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003536; AAF52931.1; -.
DR HSSP; P15086; 2HPA.
DR FlyBase; FBgn0038912; CG6656.

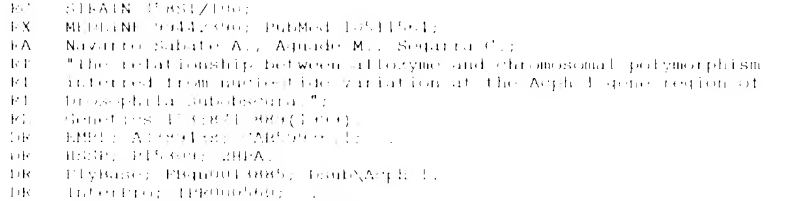


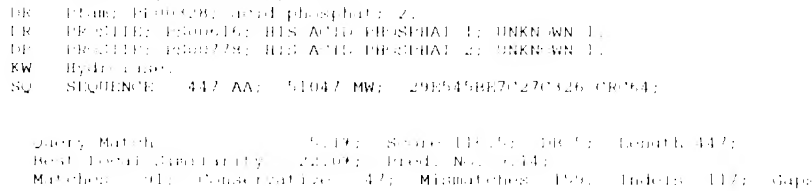
```
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   [illegible]
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003516; AAF49156.1; -.
DR   HSSP; P15309; 2HPA.
DR   FlyBase; FBgn0038752; [illegible].
DR   InterPro; IPR000560; -.
DR   Pfam; PF00328; acid_phosphat; 2.
DR   PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR   PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
SQ   SEQUENCE   392 AA;  45297 MW;  4D3F447EF6E7DCBD CRC64;

  Query Match             [illegible];  Score 124.5;  DB 5;  Length 392;
  Best Local Similarity   21.6%;  Pred. No. 0.024;
  Matches   86;  Conservative   56;  Mismatches  152;  Indels  105;  Gaps   21;

Qy         30 LESVVIVSRHGVRAPTKAT[illegible]QLMQEVTPDAW[illegible]  81

Db         36 LELLHVVFRHGPRTPA[illegible]          DTYTKDPYVNETYYPFGWGQLTNNGKEKEFNIG  84

Qy         82 [illegible]  141

Db         85 IWLRERY[illegible]  140

Qy        142 [illegible]  200

Db        141 WQPIPVESQ[illegible]KLNE[illegible]DIVLLEVEKRP[illegible]EY[illegible]FEAINPVYKEPE----  179

Qy        201 [illegible]  251

Db        180 VKAEIEPYLEMPKFIFHHPCLSFKELEDV[illegible]------------LIADLEWQK  223

Qy        252 [illegible]  311

Db        224 EIPEW---IHAYFEEKIQ[illegible]  275

Qy        312 [illegible]  367

Db        276 GILEFYCPKLIYYAIH[illegible]  [illegible]

Qy        368 [illegible]

Db        [illegible] YWVPIYP-----[illegible]  [illegible]
```

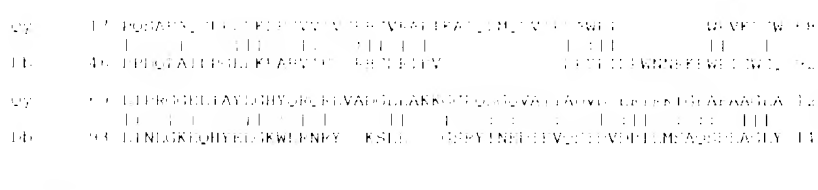
```
RESULT 7
O00092
ID   O00092      PRELIMINARY;      PRT;   465 AA.
AC   O00092;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   3-PHYTASE A PRECURSOR (EC 3.1.3.8) (MYO-INOSITOL HEXAKISPHOSPHATE
DE   PHOSPHOHYDROLASE A) (3-PHYTASE A) (MYO-INOSITOL HEXAKISPHOSPHATE
DE   PHOSPHOHYDROLASE A).
GN   PHYA.
OS   Aspergillus fumigatus (Sartorya fumigata).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5085;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 27-37.
RC   STRAIN=ATCC 34625;
RX   MEDLINE=97228967; PubMed=9143104;
RA   Pasamontes L., Haiker M., Wyss M., Tessier M., van Loon A.P.G.M.;
RT   "Gene cloning, purification, and characterization of a heat-stable
RT   phytase from the fungus Aspergillus fumigatus.";
RL   Appl. Environ. Microbiol. 63:1696-1700(1997).
CC   -!- FUNCTION: CATALYZES THE HYDROLYSIS OF INORGANIC ORTHOPHOSPHATE
CC       FROM PHYTATE.
CC   -!- CATALYTIC ACTIVITY: MYO-INOSITOL HEXAKISPHOSPHATE + H(2)O = D-MYO-
CC       INOSITOL 1,2,4,5,6-PENTAKISPHOSPHATE + PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- MISCELLANEOUS: SHOWS HIGH ACTIVITY WITHIN A NEUTRAL TO ACIDIC PH
CC       RANGE. [illegible]
CC       7.5. IT IS ABLE TO WITHSTAND TEMPERATURES UP TO 100 DEGREES
CC       CELSIUS OVER A PERIOD OF 20 MIN, WITH A LOSS OF ONLY 10% OF THE
CC       INITIAL ENZYMATIC ACTIVITY.
CC   -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
DR   EMBL; U59804; AAB96872.1; -.
DR   HSSP; P34752; 1IHP.
DR   InterPro; IPR000560; -.
DR   Pfam; PF00328; acid_phosphat; 1.
DR   PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR   PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; UNKNOWN_1.
KW   Hydrolase; Glycoprotein; Signal.
FT   SIGNAL        1     26
FT   CHAIN        27    465       3-PHYTASE A.
FT   ACT_SITE     81     81       REQUIRED FOR BINDING SUBSTRATE (BY
FT                                SIMILARITY).
FT   ACT_SITE     82     82       NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT   ACT_SITE    [illegible]    [illegible]       PROTON DONOR (BY SIMILARITY).
FT   DISULFID     30     39       BY SIMILARITY.
FT   DISULFID     70    412       BY SIMILARITY.
FT   DISULFID    213    463       BY SIMILARITY.
FT   DISULFID    [illegible]    [illegible]       BY SIMILARITY.
FT   DISULFID    434    442       BY SIMILARITY.
FT   CARBOHYD    [illegible]    [illegible]       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    119    119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    205    205       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    228    228       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    [illegible]    [illegible]       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    350    350       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    [illegible]    [illegible]       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   465 AA;  50836 MW;  86FC1D9058C9B8C9 CRC64;

  Query Match             5.2%;  Score 116;  DB 3;  Length 465;
  Best Local Similarity   21.7%;  Pred. No. 0.089;
  Matches  102;  Conservative   76;  Mismatches  183;  Indels  111;  Gaps   24;

Qy          [illegible]

Db          [illegible]

Qy          [illegible]
```

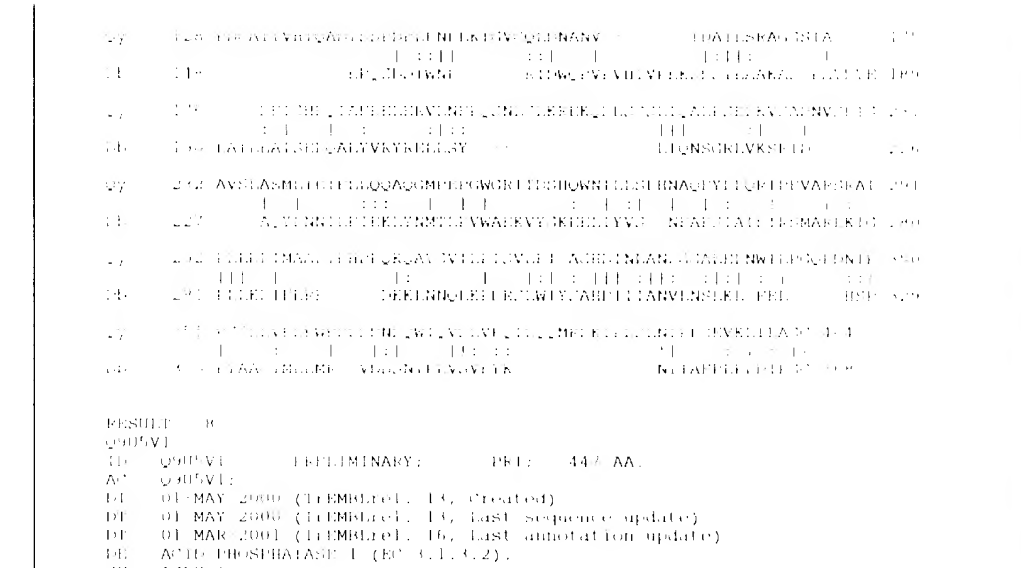


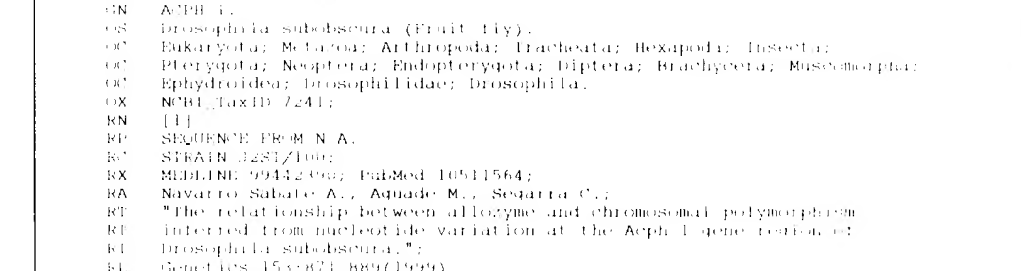


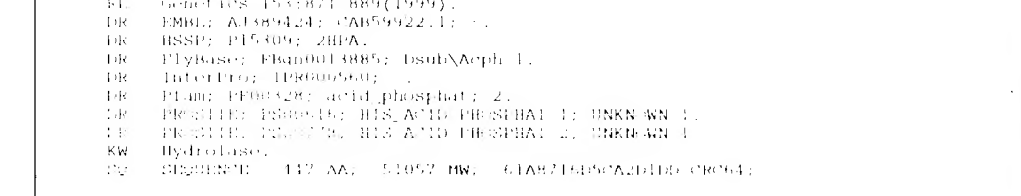


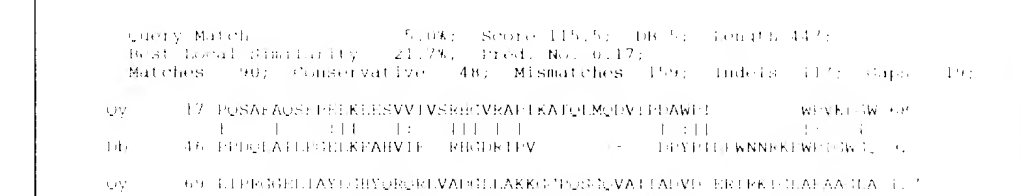


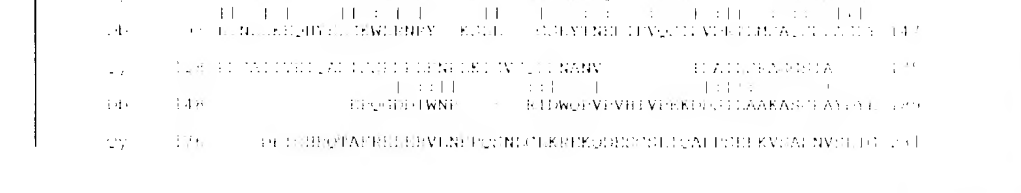


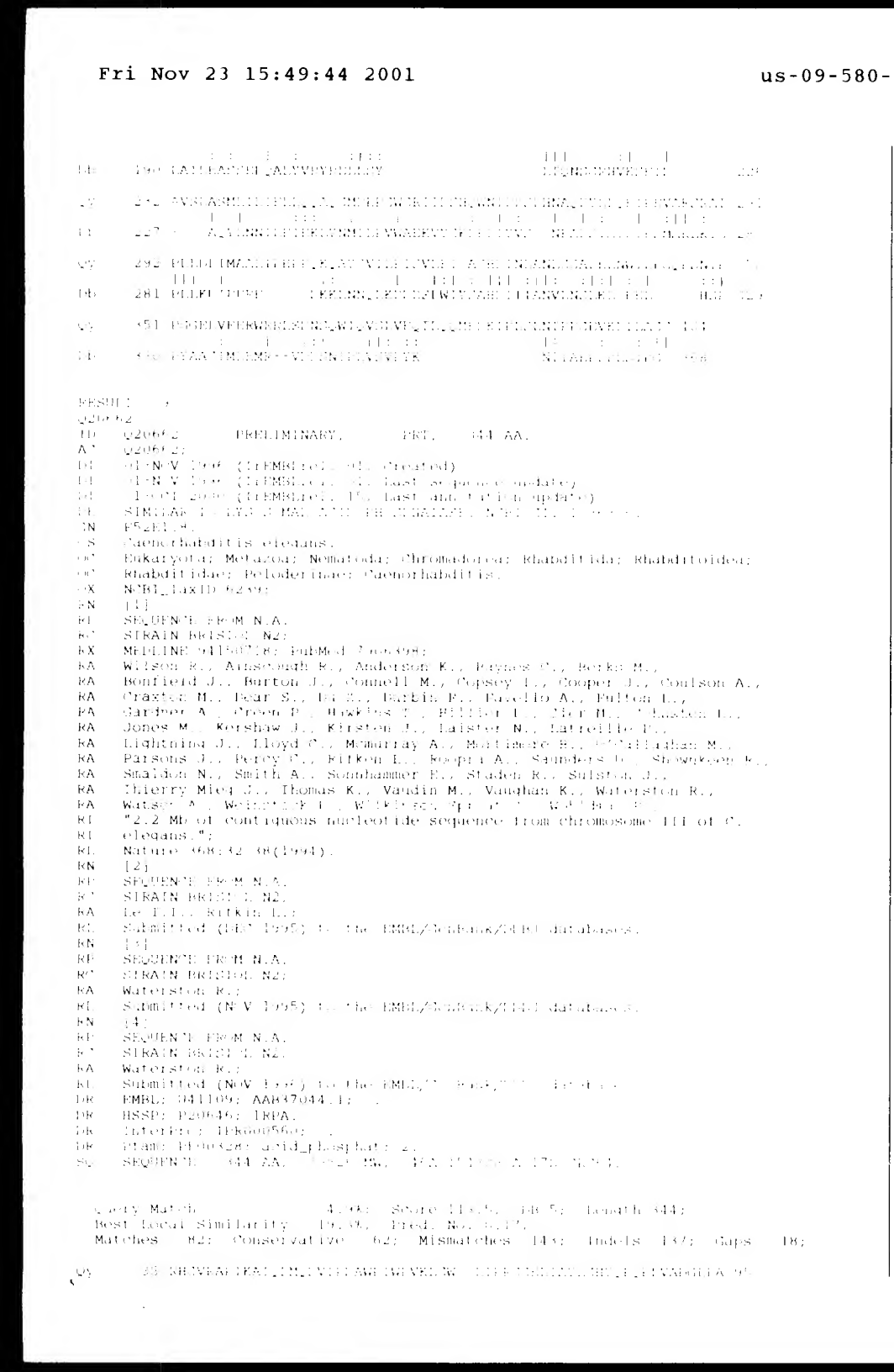






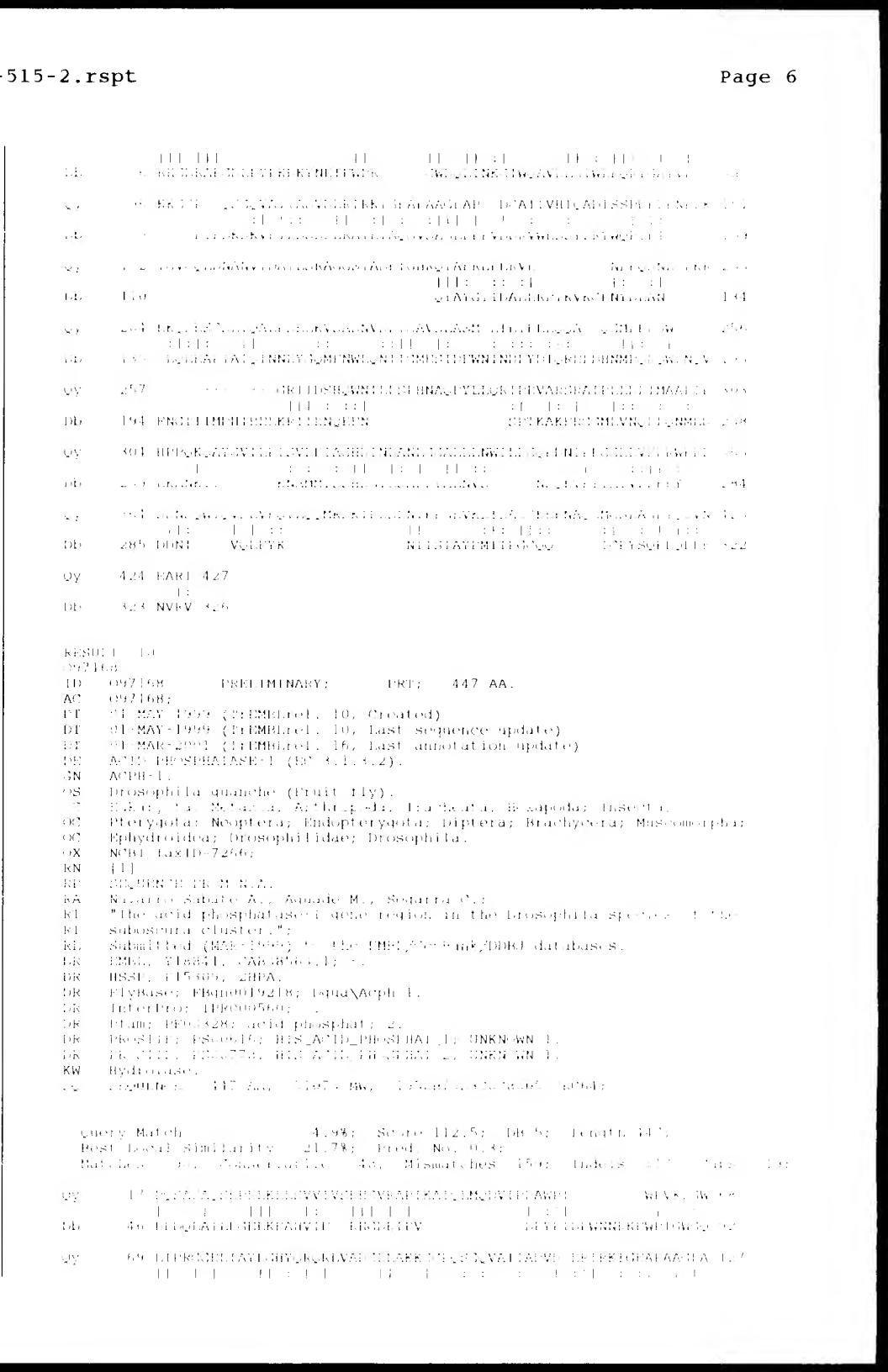






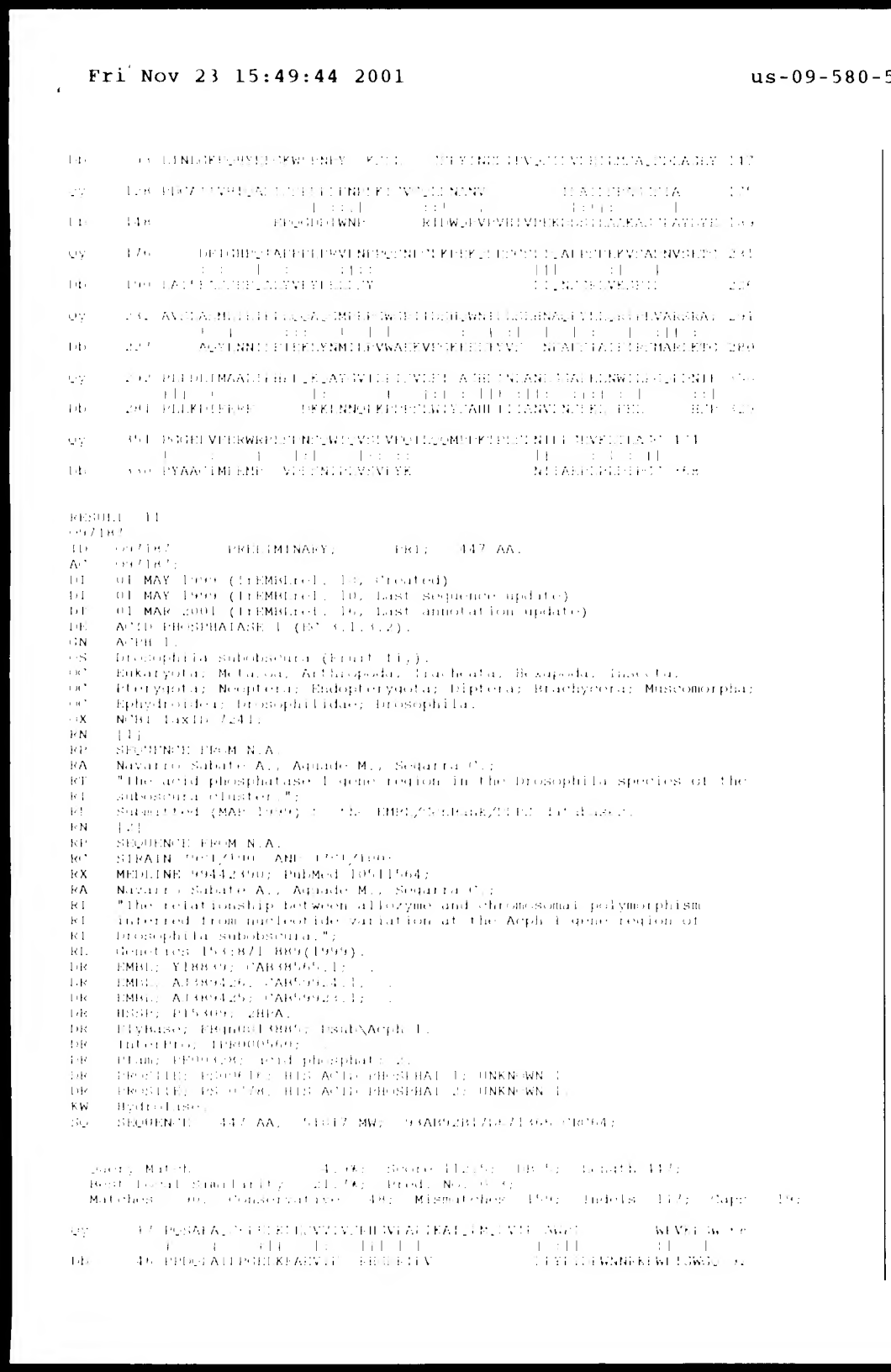
RESULT 1
Q20662
ID Q20662 PRELIMINARY; PRT; 444 AA.
AC Q20662;
DT 01-NOV-1998 (TrEMBLrel. 01, Created)
DT 01-NOV-1998 (TrEMBLrel. 01, Last sequence update)
DT 15-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SIMILAR TO LYS-2 MAL ATTE PROTEIN.
GN F52E1.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN BRISTOL N2;
RX MEDLINE=21507218; PubMed=11076968;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laisster N., Latreille P.,
RA Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN BRISTOL N2;
RA Le T.L., Kitkin L.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN BRISTOL N2;
RA Waterston R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN BRISTOL N2;
RA Waterston R.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; D41109; AAB37044.1; -.
DR HSSP; P20646; 1RPA.
DR InterPro; IPR000560; -.
DR Pfam; PF00328; acid_phosphat; 2.
SQ SEQUENCE 444 AA, [illegible] MW; [illegible] CRC32;

Query Match 4.4%; Score 112.5; DB 5; Length 444;
Best Local Similarity 19.9%; Pred. No. 6.1e-17;
Matches 82; Conservative 62; Mismatches 144; Indels 137; Gaps 18;



RESULT 10
Q97168
ID Q97168 PRELIMINARY; PRT; 447 AA.
AC Q97168;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ACPH PHOSPHATASE-1 (EC 3.1.3.2).
GN ACPH-1.
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.
RA Navarro Sabate A., Aguade M., Segarra C.;
RT "The acid phosphatase-1 gene region in the Drosophila speciose [illegible]
RT subobscura cluster.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
DR EMBL; Y16841; CAA76960.1; -.
DR HSSP; P15605; 2HPA.
DR FlyBase; FBgn0019218; Dyak\Acph-1.
DR InterPro; IPR000560; -.
DR Pfam; PF00328; acid_phosphat; 2.
DR ProSite; PS00616; HIS_ACID_PHOSPHAT_1; UNKNOWN 1.
DR PR [illegible]; PR00779; HIS_ACID_PHOSPHAT_2; UNKNOWN 1.
KW Hydrolase.
SQ SEQUENCE [illegible] AA, [illegible] MW; [illegible] CRC64;

Query Match 4.9%; Score 112.5; DB 5; Length 447;
Best Local Similarity 21.7%; Pred. No. 6.1e-17;
Matches 93; Conservative 46; Mismatches 150; Indels 127; Gaps 10;

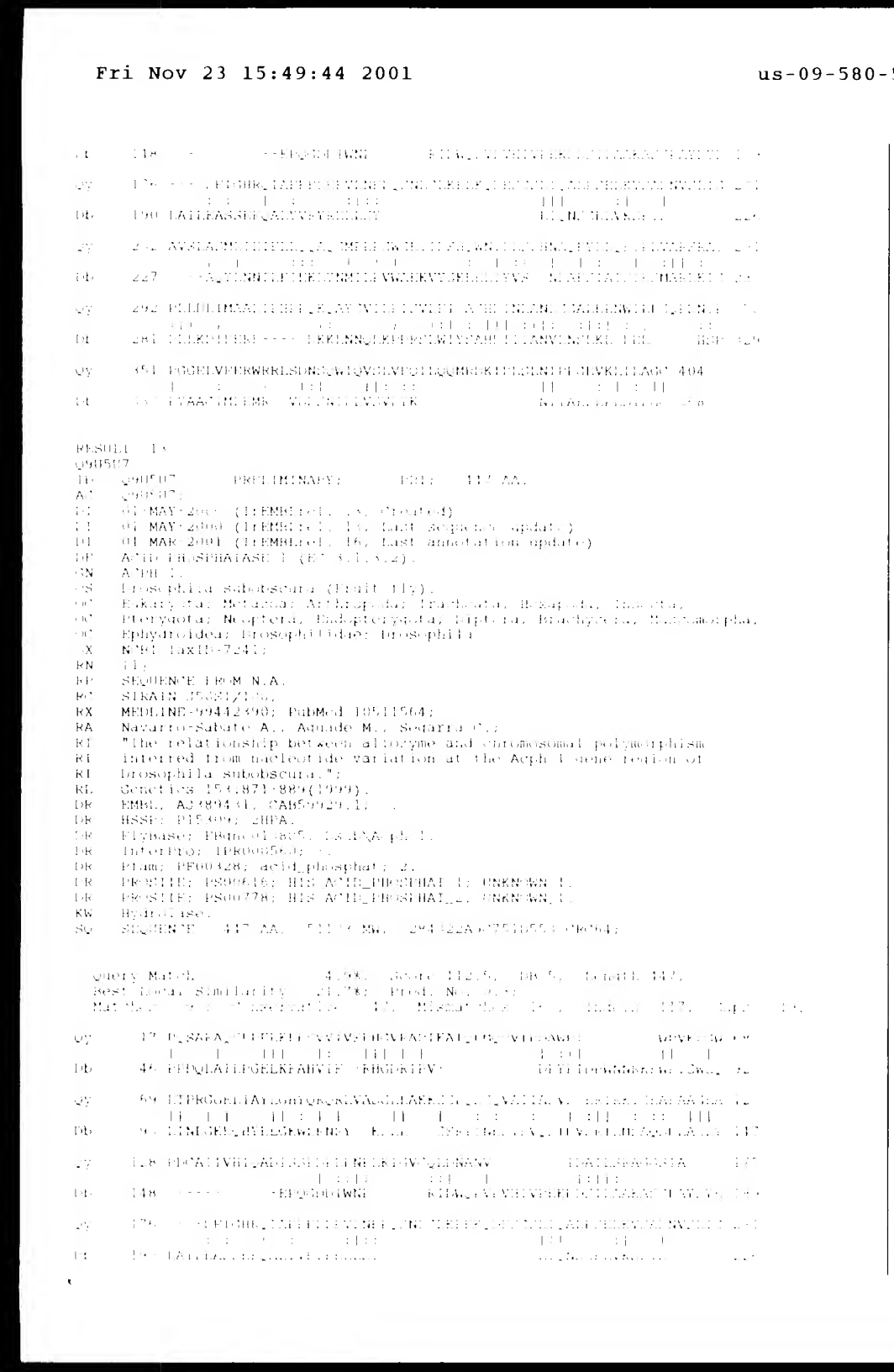


RESULT 11
Q97187
ID Q97187 PRELIMINARY; PRT; 447 AA.
AC Q97187;
DT 01-MAY-1999 (TrEMBLrel. 13, Created)
DT 01-MAY-1999 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ACID PHOSPHATASE 1 (EC 3.1.3.2).
GN ACPH 1.
OS Drosophila subobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7241;
RN [1]
RP SEQUENCE FROM N.A.
RA Navarro-Sabate A., Aquade M., Segarra C.;
RT "The acid phosphatase 1 gene region in the Drosophila species of the
RT subobscura cluster.";
RL Submitted (MAR-2000) to EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN "ORI/2000, ANE-1221/2000";
RX MEDLINE 99442390; PubMed-10511564;
RA Navarro-Sabate A., Aquade M., Segarra C.;
RT "The relationship between allozyme and chromosomal polymorphism
RT inferred from nucleotide variation at the Acph-1 gene region of
RT Drosophila subobscura.";
RL Genetics 153:871-880(1999).
DR EMBL; Y18849; CAB59927.1; -.
DR EMBL; AJ309429; CAB59927.1; JOINED.
DR EMBL; AJ309426; CAB59927.1; JOINED.
DR EMBL; AJ309429; CAB59927.1; JOINED.
DR HSSP; P15309; 2BPA.
DR FlyBase; FBgn0011885; Dsub\Acph-1.
DR InterPro; IPR000560; -.
DR Pfam; PF00328; acid_phosphat; 2.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; UNKNOWN 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; UNKNOWN 1.
KW Hydrolase;
SQ SEQUENCE 447 AA; 51017 MW; 99AB928B7A871965 CRC64;

Query Match 4.9%; Score 112.5; DB 5; Length 447;
Best Local Similarity 21.7%; Pred. No. 0.5;
Matches 90; Conservative 48; Mismatches 159; Indels 117; Gaps 19;

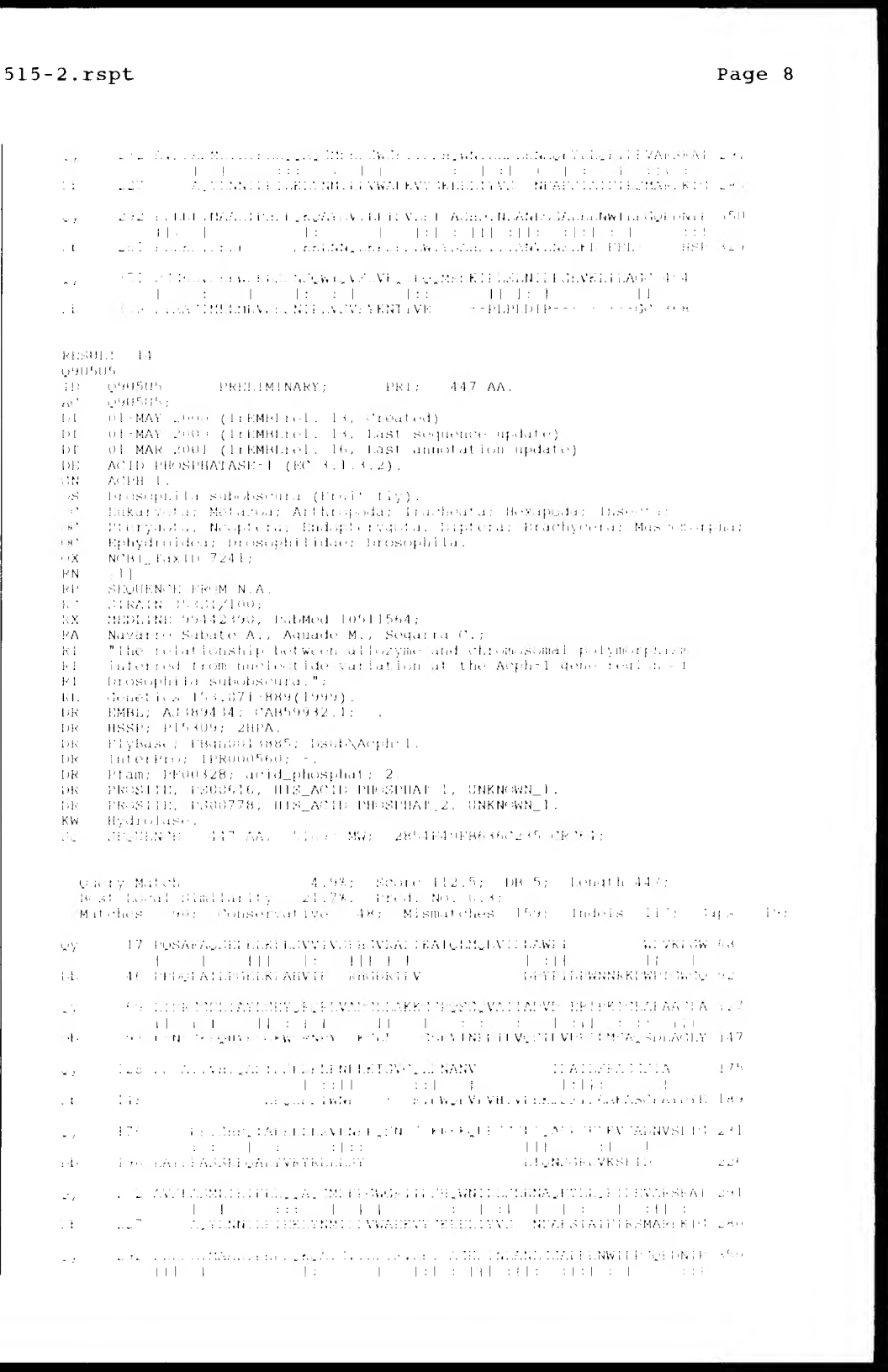
RESULT 12
Q98U89
ID Q98U89 PRELIMINARY; PRT; 447 AA.
AC Q98U89;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ACID PHOSPHATASE-1 (EC 3.1.3.2).
GN ACPH-1.
OS Drosophila subobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7241;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN JL2S1/108;
RX MEDLINE 99442390; PubMed-10511564;
RA Navarro-Sabate A., Aquade M., Segarra C.;
RT "The relationship between allozyme and chromosomal polymorphism
RT inferred from nucleotide variation at the Acph-1 gene region of
RT Drosophila subobscura.";
RL Genetics 153:871-880(1999).
DR EMBL; AJ309429; CAB59927.1; -.
DR HSSP; P15309; 2BPA.
DR FlyBase; FBgn0011885; Dsub\Acph-1.
DR InterPro; IPR000560; -.
DR Pfam; PF00328; acid_phosphat; 2.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; UNKNOWN 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; UNKNOWN 1.
KW Hydrolase;
SQ SEQUENCE 447 AA; 51074 MW; 8B21FB781B6BDA09 CRC64;

Query Match 4.9%; Score 112.5; DB 5; Length 447;
Best Local Similarity 21.7%; Pred. No. 0.5;
Matches 90; Conservative 48; Mismatches 159; Indels 117; Gaps 19;



RESULT 13
Q90507
ID Q90507 PRELIMINARY; PRT: 447 AA.
AC Q90507;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ACID PHOSPHATASE-1 (EC 3.1.3.2).
GN ACPH 1.
OS Drosophila subobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7241;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN 25631/100;
RX MEDLINE=99442390; PubMed=10511564;
RA Navarro-Sabate A., Aquade M., Segarra C.;
RT "The relationship between allozyme and chromosomal polymorphism
RT inferred from nucleotide variation at the Acph-1 gene region of
RT Drosophila subobscura.";
RL Genetics 153:871-889(1999).
DR EMBL; AJ389434; CAB59929.1; -.
DR HSSP; P15309; 2HPA.
DR Flybase; FBgn0013885; DsubAcph-1.
DR InterPro; IPR000560; -.
DR Pfam; PF00328; acid_phosphat; 2.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; UNKNOWN_1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 447 AA; 51174 MW; 2941E2A2F751D55C CRC64;

Query Match 4.9%; Score 112.5; DB 5; Length 447;
Best Local Similarity 21.7%; Pred. No. 6.5;
Matches 96; Conservative 49; Mismatches 159; Indels 117; Gaps 15;



RESULT 14
Q90505
ID Q90505 PRELIMINARY; PRT: 447 AA.
AC Q90505;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ACID PHOSPHATASE-1 (EC 3.1.3.2).
GN ACPH 1.
OS Drosophila subobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7241;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN 25631/100;
RX MEDLINE=99442396; PubMed=10511564;
RA Navarro-Sabate A., Aquade M., Segarra C.;
RT "The relationship between allozyme and chromosomal polymorphism
RT inferred from nucleotide variation at the Acph-1 gene region of
RT Drosophila subobscura.";
RL Genetics 153:871-889(1999).
DR EMBL; AJ389434; CAB59932.1; -.
DR HSSP; P15309; 2HPA.
DR Flybase; FBgn0013885; DsubAcph-1.
DR InterPro; IPR000560; -.
DR Pfam; PF00328; acid_phosphat; 2.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; UNKNOWN_1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 447 AA; 51174 MW; 2941E2A2F751D55C CRC64;

Query Match 4.9%; Score 112.5; DB 5; Length 447;
Best Local Similarity 21.7%; Pred. No. 6.5;
Matches 96; Conservative 49; Mismatches 159; Indels 117; Gaps 15;

```
Db      291 PLIKDIIIEKI      DKKLNNQLKPIRSEWIYSAERIIIANVLNSLKL PEL      HSP  329

Qy      351 PGSDELVPERWPPLSDNSQWTQVSLVPQTLQQMRDKIPLSLNTIPPSEVKLLLAQC 404
            |   :   |    |:|    ||: ::             ||    : | : ||
Db      330 PYAACIMLLMK  VDLSNIPLVSVLYK             NIIAEILPLLIPSC 368


RESULT  2
Q00502
ID   Q00502       PRELIMINARY;      PRT;   447 AA.
AC   Q00502;
DT   01-MAY-1990 (IntEMBLrel. 13, Created)
DT   01-MAY-2000 (IntEMBLrel. 63, Last sequence update)
DT   01-MAY-2001 (IntEMBLrel. 65, Last annotation update)
DE   ACPH-1 PHOSPHATASE 1 (EC 3.1.4.-).
GN   ACPH-1.
OS   Drosophila subobscura (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7241;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN: Poos/1969;
RX   MEDLINE=90144390; PubMed=1970564;
RA   Navarro-Sabate A., Aguade M., Segarra C.;
RT   "The relationship between allozyme and chromosomal polymorphism
RT   inferred from nucleotide variation at the Acph-1 locus region of
RT   Drosophila subobscura.";
RL   Genetics 124:921-929(1990).
DR   EMBL; A35044; CAA78644.1; -.
DR   HSSP; P15309; 2HPA.
DR   FlyBase; FBgn0011088; Gsub\Acph-1.
DR   InterPro; IPR000560; -.
DR   Pfam; PF00328; acid phosphat; 2.
DR   PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; UNKNOWN_1.
DR   PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; UNKNOWN_1.
KW   Hydrolase.
SQ   SEQUENCE   447 AA;  51108 MW;  6ADF6BEAA4E9F4B0 CRC64;


  Query Match             4.7%;  Score 112.5;  DB 5;  Length 447;
  Best Local Similarity   21.7%;  Pred. No. 0.5;
  Matches   90;  Conservative   48;  Mismatches  159;  Indels  117;  Gaps   19;

Qy       17 PQSAFAQSEPELKLESVVISPRGVRAPTKAIGIMQIVIPLAWPE    --WPVKLGW 68
            |   |   |||   |:   ||| | |             | :||           ||   |
Db       46 PEDQLAILPGELKFAHVIP  PRGSPLPV            EPYSIDEWNNRKPWPLSGWGQ 92

Qy       69 LIPGGIPLIAYLGHYSHYQSQKLVAVSLLAKKGQIQGSQVAIIALVD EKLPKLGLAIAALA 127
            ||  | |    ||:|| | |     ||   |   : :   :   | :||  : ::  |||
Db       93 LINQSKLQHYLPSKWLENPY  KSCLL  QSRYINRLLLVQSLLVLGKLLMSAQSDLAGLY 147

Qy      128 PDCAILVHLQADISSPDQLENPLKLGVQLLDNANV          LDAILSPAGGSIA--  175
                    | :::|     ::|  :           |:|||      |
Db      148             ELQSDLIWNP    KDIWQFVLVHLVLVGKDASLLAAKAASLATAYE 189

Qy      176      DFTGHPQLAFLELLDVLNFLQSNLQLKPFKQDESSLLQALESELKVSADNVSLIG 231
                 .   | :  :|::          |||   |  |
Db      190 LATLEASSEPQALYVKYPELLSY          LLQNSGDLVKSFID        226

Qy      232 AVGLAAMLTLGLLLQLAQMLFGWGFTLSHQWNLLGNLNAQYLGLQFLVAPGRAL 291
                | |     :::   |  | |      :  | :|  |  | :   |  :|| :
Db      227      AQVLNNLGSFGKLNMLLVWALKVYGELLYYG  NLALSIALLGSMAKLKLG 280

Qy      292 PLLDLIMAALLLHFQGQAYSVLLPLGVLLL AGHVLNLANLSSAFLLNWLLSGLLNLP 350
            |||  |           |:      |   |:| : ||| :||:  :|,  : )      :|
Db      281 PLLKDLLLLKL      DKKLNNQLKPIRSEWIYSAERIIIANVLNSLKL PEL      HSP 329

Qy      351 PGSDELVPERWPPLSDNSQWTQVSLVPQTLQQMRDKIPLSLNTIPPSEVKLLLAQC 404
            |   :   |    |:|    ||: ::             ||    : | : ||
Db      330 PYAACIMLLMK  VDLSNIPLVSVLYK             NIIAEILPLLIPSC 368


Search completed: November 20, 2001, 10:03:49
Job time: 181 sec
```

SM protein - protein search, using sw model

Run on: November 20, 2001, 10:06:13 ; Search Time 13.13 Seconds
(without alignments)
1147.937 Million cell updates/sec

Title: US-09-580-515-2
Perfect score: 2235
Sequence: 1 MKAILIPFLSLLIPLLIQSA..... ...IVNEARTPAVSERSHHHHHH 440

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93457 seqs, 34550496 residues

Total number of hits satisfying chosen parameters: 93457

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt 39:*

Prob. N.: is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

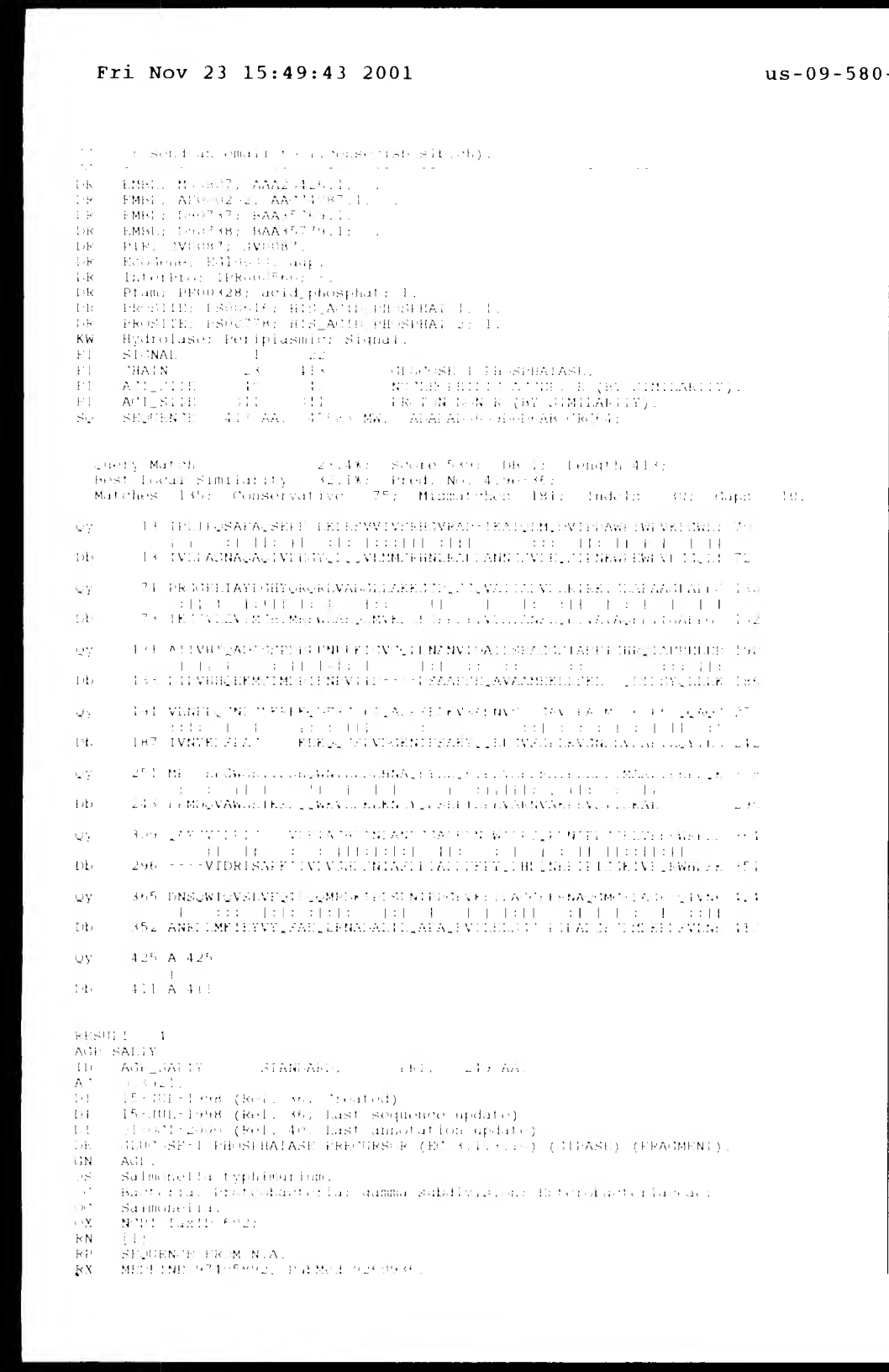
Result No.	Score	Query Match	Length	DB	ID	Description
1	2235	97.1	442	1	PPA_ECOLI	P07102 escherichia
2	541	23.5	417	1	AGP_PRORE	Q52309 providencia
3	539	23.4	413	1	AGP_ECOLI	P19926 escherichia
4	279	12.1	389	1	AGP_SALTY	Q7CQ21 salmonella
5	139.5	6.1	428	1	PPAL_HUMAN	P11117 homo sapien
6	141	5.7	423	1	PPAL_RAT	P20611 rattus norv
7	122.5	5.3	421	1	PPAL_MOUSE	P49638 mus musculu
8	117	5.1	466	1	PPAP_HUMAN	P15309 homo sapien
9	109.5	4.8	457	1	PAX8_MOUSE	Q00288 mus musculu
10	109	4.7	2054	1	ZLES_DROME	P13368 drosophila
11	107	4.6	450	1	PXRA_HUMAN	Q96716 homo sapien
12	106.5	4.6	1048	1	SUCC_ECOLI	P13458 escherichia
13	105.5	4.7	459	1	PXRA_CANFA	P47240 canis famili
14	102	4.4	638	1	SRPK_HUMAN	P98240 homo sapien
15	101.5	4.4	479	1	PHYB_ASPAW	P34755 aspergillus
16	101	4.4	458	1	PAX8_RAT	P51034 rattus norv
17	100.5	4.4	469	1	HBM3_PEA	Q43092 pisum sativ
18	100	4.3	1354	1	VCAP_HSVI1	P06491 herpes simp
19	97	4.2	587	1	DCDA_ECOLI	P31838 escherichia
20	97	4.2	4,289	1	TENX_HUMAN	P22105 homo sapien
21	96.5	4.2	463	1	PPAP_RAT	P20648 rattus norv
22	96.5	4.2	2695	1	HMI2_YEREN	P38933 yersinia en
23	95.5	4.1	4,994	1	PCHM_HUMAN	P98160 homo sapien
24	9	4.1	1251	1	YQ1D_CAEEL	Q09550 caenorhabdi
25	94.5	4.1	474	1	PHYB_ASPNG	P34754 aspergillus
26	93.5	4.1	8828	1	TRX_DROVI	Q24742 drosophila
27	93	4.2	776	1	HYPE_AZOCH	Q44950 azotobacter
28	92	4.0	8043	1	POLG_HCV36	P26660 h genome po
29	91.5	4.0	792	1	HEXX_ADE12	P03712 human adeno
30	91.5	4.0	635	1	TPOR_HUMAN	P40238 homo sapien
31	91	4.0	625	1	TPOR_MOUSE	Q08351 mus musculu
32	91	4.0	648	1	SRPK_CANFA	P06625 canis famil
33	90.5	4.0	1085	1	CASK_BOVIN	P35384 bos taurus
34	90.5	4.0	1156	1	HIS_DROME	Q01147 drosophila
35	90	4.0	455	1	VL2_HSVI1	P14338 human herpe
36	90	4.0	921	1	120K_DROME	P17825 drosophila
37	89.5	3.9	618	1	ZA36_MOUSE	P47804 mus musculu
38	89.5	3.9	847	1	NCM2_MOUSE	Q03149 mus musculu
39	89.5	3.9	1589	1	PHP_DROME	P39269 drosophila
40	89	3.9	411	1	PPAY_CAEEL	Q10644 caen rhabdi
41	89	3.9	484	1	PEPD_ECOLI	P15288 escherichia
42	89	3.9	809	1	QUIA_ACIAD	Q59080 acinetobact
43	89	3.9	811	1	PRIA_RHERU	P55445 rhodospiril
44	89	3.9	971	1	LONM_CAEEL	Q41952 caenorhabdi
45	88.5	3.8	1174	1	NIPS_ECOLI	P15264 escherichia

ALIGNMENTS

RESULT 1
PPA_ECOLI
ID PPA_ECOLI STANDARD; PRT; 442 AA.
AC P07102;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PERIPLASMIC APPA PROTEIN PRECURSOR [INCLUDES: PHOSPHOHYDROLASE
DE PHOSPHOHYDROLASE (EC 3.1.3.2) (PH 2.5 ACID PHOSPHATASE) (AE)
DE PHYTASE (EC 3.1.3.26)].
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-29.
RC STRAIN=K12;
RX MEDLINE=90306615; PubMed=2168385;
RA Dassa J., Marck C., Boquet P.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase.";
RL J. Bacteriol. 172:5497-5500(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97504232; PubMed=9005232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto S.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 1-112 FROM N.A.
RX MEDLINE=87271766; PubMed=3038201;
RA Touati E., Danchin A.;
RT "The structure of the promoter and amino-terminal region of the pH
RT 2.5 acid phosphatase structural gene (appA) of E. coli: a negative
RT control of transcription mediated by cyclic AMP.";
RL Biochimie 69:215-221(1987).
RN [5]

```
SQ   SEQUENCE   [illegible] MW;  [illegible] CRC64;
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE; [illegible].
RA   Greiner R., Konietzny U., Jany K.-D.;
RT   "Purification and characterization of two phytases from Escherichia
RT   coli.";
RL   Arch. Biochem. Biophys. 303:107-113(1993).
RN   [6]
RP   CHARACTERIZATION AND SEQUENCE OF 23-34.
RA   Greiner R., Jany K.-D.;
RT   "Characterization of a phytase from Escherichia coli.";
RL   Biol. Chem. Hoppe-Seyler 373:657-663(1992).
RN   [7]
RP   CHARACTERIZATION AND SEQUENCE OF 23-34.
RX   MEDLINE; [illegible]; PubMed [illegible].
RA   Greiner R., Konietzny U., Jany K.-D.;
RT   "Purification and characterization of two phytases from Escherichia
RT   coli.";
RL   Arch. Biochem. Biophys. 303:107-113(1993).
RN   [8]
RP   MUTAGENESIS.
RX   MEDLINE=96054596; PubMed=8629961;
RA   Golovan S., Wang G., Zhang J., Forsberg C.W.;
RT   "Characterization and overproduction of the Escherichia coli appA
RT   encoded bifunctional enzyme that exhibits both phytase and acid
RT   phosphatase activities.";
RL   Can. J. Microbiol. 46:59-71(2000).
RN   [9]
RP   MUTAGENESIS.
RX   MEDLINE=96054596; PubMed=8629961;
RA   Golovan S., Wang G., Zhang J., Forsberg C.W.;
RT   "Characterization and overproduction of the Escherichia coli appA
RT   encoded bifunctional enzyme that exhibits both phytase and acid
RT   phosphatase activities.";
RL   Can. J. Microbiol. 46:59-71(2000).
RN   [10]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE=20122624; PubMed=10556111;
RA   Lim D., Golovan S., Forsberg C.W., Jia Z.;
RT   "Crystal structures of Escherichia coli phytase and its complex with
RT   phytate.";
RL   Nat. Struct. Biol. 7:108-113(2000).
CC   -!- CATALYTIC ACTIVITY: AN 8-PHYTANYL-6-O-(2,
CC       ALPHA-D-GLUCOPYRANOSYL)-D-MYO-INOSITOL 1,2,3-
CC   -!- CATALYTIC ACTIVITY: MYO-INOSITOL HEXAKISPHOSPHATE + H(2)O =
CC       1-MYO-INOSITOL 1,2,3,4,5-PENTAKISPHOSPHATE + ORTHOPHOSPHATE.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC   -!- INDUCTION: IN ANERIT N IS CAMP-MEDIATED CONTROL; THIS ENZYME IS
CC       INDUCED WHEN BACTERIAL CELLS ENTER STATIONARY PHASE; ITS
CC       SYNTHESIS IS STIMULATED BY GLUCOSE AND REPRESSED BY
CC       AEROBIOSIS AND ANAEROBIOSIS CONDITIONS.
CC   -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC   ---------- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ----------
DR   EMBL; M58708; AAA72086.1; -.
DR   EMBL; AE000200; AAC74065.1; -.
DR   EMBL; [illegible]; AAA23745.1; -.
DR   EMBL; X05471; CAA29031.1; -.
DR   EMBL; M58708; -; NOT_ANNOTATED_CDS.
DR   EMBL; [illegible]; -; -.
DR   PIR; A26584; A26584.
DR   PIR; B36774; B36774.
DR   PIR; S17960; S17960.
DR   PIR; S18018; S18018.
DR   PIR; [illegible]; [illegible].
DR   PIR; [illegible]; [illegible].
DR   PDB; 1DKI; 03-AUG-00.
DR   PDB; 1DKJ; 03-AUG-00.
DR   PDB; 1DKN; 03-AUG-00.
DR   PDB; 1DKO; 03-AUG-00.
DR   PDB; 1DKP; 03-AUG-00.
DR   PDB; 1DKQ; 03-AUG-00.
DR   SWISS-2DPAGE; P07102; ECOLI.
DR   EcoGene; EG10049; appA.
DR   InterPro; IPR000560; -.
DR   Pfam; PF00328; acid_phosphat; 1.
DR   PROSITE; PS00616; HTS_ACID_PHOSPHAT_1; 1.
DR   PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW   Hydrolase; Periplasmic; Signal; Multifunctional enzyme; [illegible].
FT   SIGNAL       1     22
FT   CHAIN       23    432       [illegible]
FT   ACT_SITE    59     59       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    62     62       PROTON DONOR.
FT   DISULFID    99    130
FT   DISULFID   155    430
FT   DISULFID   200    210
FT   DISULFID   404    413
FT   [illegible]  [illegible]  [illegible]  [illegible]
FT                                REF. 4).
FT   CONFLICT     25     26       EL -> DV (IN REF. 4).
FT   CONFLICT    112    112       D -> S (IN REF. 4).
SQ   SEQUENCE   432 AA;  47859 MW;  [illegible] CRC64;


  Query Match          97.1%;  Score 2295;  DB 1;  Length 432;
  Best Local Similarity  99.5%;  Pred. No. 1.6e-172;
  Matches  [illegible];  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy       1 MKAILIPFFSALICPLCQAHAAQTEPELESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKAILIPFFSALICPLCQAHAAQSEPELESVVIVSRHGVRAPTKATQLMQDVTPDAWP 60

Qy      61 AWPVKLGWLTPRGGEMIALMGGFYQRQKFQSQGLLAKGCPTPNSIYVWADVDERTRKT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AWPVKLGWLTPRGGEMIALMGGFYQRQKFQSQGLLAKGCPTPNSIYVWADVDERTRKT 120

Qy     121 AEAAAAAGLAPGCAITVHTQADEKSSDPLFNPLKTGVCQLDNANVTDAILSKAGGSIADFT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AEAAAAAGLAPGCAITVHTQADEKSSDPLFNPLKTGVCQLDNANVTDAILSKAGGSIADFT 180

Qy     181 GFKTAFGNMYLNTEDKNSQVFQSKYNCLSDHAPVSARVDPSFQETKYLSMVDKSNGAGGE 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GFKTAFGNMYLNTEDKNSQVFQSKYNCLSDHAPVSARVDPSFQETKYLSMVDKSNGAGGE 240

Qy     241 YLVMLVSAHLEPVSMTTGPDVKLTSSHAVSVTVGAKWSPYLTAKIVPLAIQTRNTPKVQF 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 YLVMLVSAHLEPVSMTTGPDVKLTSSHAVSVTVGAKWSPYLTAKIVPLAIQTRNTPKVQF 300

Qy     301 TQTVADQSTQEVYSHLTKAVRFAMTDIEAGVPDKLRPLFGLVKTLDMSDNFGGLSKPGAV 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TQTVADQSTQEVYSHLTKAVRFAMTDIEAGVPDKLRPLFGLVKTLDMSDNFGGLSKPGAV 360

Qy     361 TVNEARIPAGSLETF 420
           |||||||||||||||
Db     361 TVNEARIPAGSLETF 420


RESULT  2
ACP_PROPE
ID   [illegible]  [illegible]  [illegible]  [illegible] AA.
AC   Q52309;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
```

```
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   GLUCOSE 1 PHOSPHATASE PRECURSOR (EC 3.1.3.10) (G1PASE).
GN   AGP.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN K12;
RA   Rocco M.L., Chiocchini A., Lombardi G., Satta G.;
RL   Submitted (SEP-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: D-GLUCOSE 1-PHOSPHATE + H(2)O = D-GLUCOSE +
CC       ORTHOPHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: PERIPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60621; CAA42900.1; -.
DR   InterPro; IPR000560; -.
DR   Pfam; PF00328; acid_phosphat; 1.
DR   PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR   PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW   Hydrolase; Periplasmic; Signal.
FT   SIGNAL       1     23       POTENTIAL.
FT   CHAIN       24    417       GLUCOSE 1 PHOSPHATASE.
FT   ACT_SITE    42     42       NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT   ACT_SITE   414    414       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   417 AA;  46940 MW;  8CC4CA4F5C18FB00 CRC64;


  Query Match             23.5%;  Score 541;  DB 1;  Length 417;
  Best Local Similarity   30.5%;  Pred. No. 5.4e-36;
  Matches  129;  Conservative   84;  Mismatches  195;  Indels   29;  Gaps    9;

Qy      10 SELIPLITQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWPVWPVKLGWL 69
           ||  | | |  |  ||| | | |||||| | |  | | ||  |||| | |||
Db      19 ALFAPIAPIHADPTVNAGMVQVQVTKSGRNPLDFLVNETTYPVIDTAKRGIFMQYKQLGAL 78

Qy      70 TPKGQELIAYLGHYQRQRLVADGLLAKKGCPQSGQ VAIIADVDERTRKTGEAFAAGLAP 128
           | | | |  |||| | |  | ||| | || | | |  | | | |  |||| |
Db      79 LQQGALPFVSMIPVEDSKD LNKLAALLTLHNQEAKPLLQLAKQIAPESATAAVMGASF 138

Qy     129 DCATIVELQAIRSGEDPLFNPLKTGVQLANGVDTAATSAFLTAGLIG HEQTAFR 186
           |  | | |  |||||| |  |  | |  | | | | |  | | |  | |||
Db     139 GTKVNLHHQPETGDRMSVEASTL  NSTLTLMDLRAAAMGKLKPGEKSAYTL 194

Qy     187 ELERVLNFESKLCNGCKKEKQSEESNSSLALPSELAKVSADNVSLIGAVGLAMLLEGTLLP 246
           || ||  |  |    | |  | ||  |  |  |  |  |  |  |  |  |  ||
Db     195 ELDIVLNFPSKLPD  KTPKLSNETSVLNGSLLACPLEVSDSLFLSNCAVDAATAL 249

Qy     247 QAQGMP  EPGWGKLTDSHQWNTLLSLHNAQFYLQGIPEVARSRATPLLDLIKAALTPH 304
           |  | |  |  | | | | | | | | | | | | | | | | | | | | | | |
Db     250 YYEGTPAIQVAWGLVDLTEPFWFKLNTLENAVKQTPLSPTPVDAENSVAHPFNSYIDPGFVSV 309

Qy     305 TPQRQAYTVTLPESVLFTAHDTNLANLLSALGLFNGVWPPFASNLMVELYQHKNGQEFVRV 364
           |  | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     310 DKCDTA  PFTDVSLDEINAKLMKAPMKKTPLSLNTPSGEVKLTLAPL ENNAQPMSLAGIQV 369

Qy     365 SLNCQQWTQVCQLVPLQGAMHKKTPLSLNTPSGEVKLTLAPL ENNAQPMSLAGIQV 422
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     370 SIKKDDMKVFFVVSLTAPQGTPENAYLSEIPPKGDVTDDMTVEEN TPVTWSWPFQKVM 429

Qy     423 NEA 425
           |||
Db     430 AKA 432


RESULT   3
ASP_ECOLI
ID   ASP_ECOLI       STANDARD;       PRT;   413 AA.
AC   P19926;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GLUCOSE 1-PHOSPHATASE PRECURSOR (EC 3.1.3.10) (G1PASE).
GN   AGP.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-31.
RC   STRAIN K12;
RX   MEDLINE=90160318; PubMed=2153660;
RA   Pradel E., Marck C., Boquet P.L.;
RT   "Nucleotide sequence and transcriptional analysis of the Escherichia
RT   coli uup gene encoding periplasmic acid glucose 1-phosphatase.";
RL   J. Bacteriol. 172:2-7(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN K12 / MG1655;
RX   MEDLINE=97420617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto S.-K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   SEQUENCE OF 23-34.
RC   STRAIN K12 / MG1655;
RX   MEDLINE=97445075; PubMed=9290664;
RA   Link A.J., Robison K., Church G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded in
RT   the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
CC   -!- FUNCTION: ARE SUITABLY REQUIRED FOR THE GROWTH OF E.COLI IN A BETA-
CC       GLYCEROL PHOSPHATE MEDIUM CONTAINING G-3-P AS THE SOLE SOURCE OF
CC       PHOSPHATE.
CC   -!- CATALYTIC ACTIVITY: D-GLUCOSE 1-PHOSPHATE + H(2)O = D-GLUCOSE +
CC       ORTHOPHOSPHATE.
CC   -!- ENZYME REGULATION: INDEPENDENT FROM INORGANIC PHOSPHATE
CC       AVAILABILITY, AND APPARENTLY SUBMITTED TO CATABOLITE REPRESSION.
CC       IT IS POSITIVELY CONTROLLED BY CAMP AND THE CAMP RECEPTOR PROTEIN.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC   -!- MISCELLANEOUS: OPTIMALLY ACTIVE AROUND PH 4.
CC   -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```


Query Match 25.4%; Score 570; DB 2; Length 414;
Best Local Similarity 52.1%; Pred. No. 4.3e-46;
Matches 136; Conservative 75; Mismatches 101; Indels 40; Gaps 10;
Query Match 25.4%; Score 570; DB 2; Length 414;
Best Local Similarity 52.1%; Pred. No. 4.3e-46;
Matches 136; Conservative 75; Mismatches 101; Indels 40; Gaps 10;
RESULT 1
AGE_SALTY
ID AGE_SALTY STANDARD; PRT; 424 AA.
AC P11117;
DT 01-JUL-1989 (Rel. 07, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GLUCOSE-1 PHOSPHATASE PRECURSOR (EC 3.1.3.10) (G1PASE) (FRAGMENT).
GN AGE.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97105992; PubMed=9020944;
RA Gupta S.L., Wu H.C., Rick P.D.;
RC "A Salmonella typhimurium genetic locus which confers resistance
RC to colicin on copper sensitive mutants of Escherichia coli.";
RL J. Bacteriol. 179:4977-4984(1997).
CC -!- CATALYTIC ACTIVITY: D-GLUCOSE 1-PHOSPHATE + H(2)O = D-GLUCOSE +
CC ORTHOPHOSPHATE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; U27594; AAC45001.1; -.
DR StyGene; SG10595; asp.
DR InterPro; IPR000560; -.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; FAKE1AL.
KW Hydrolase; Periplasmic; Signal.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 424 GLUCOSE-1-PHOSPHATASE.
FT ACT_SITE 40 40 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 249 249 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 424 AA; 47136 MW; F8B9125BE612F7BE CRC64;
RESULT 2
PPAI_HUMAN
ID PPAI_HUMAN STANDARD; PRT; 424 AA.
AC P11117;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LYSOSOMAL ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2) (LAP).
GN ACP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=89095245; PubMed=3191910;
RA Rohlmann A., Krentler C., Schmidt B., Schroer W., Gieselmann V.;
RA Zeller C., Meisemann S., Suter U., Waheed A., Gieselmann S.;
RA Klein E.H., Neskich A., von Figura K.;

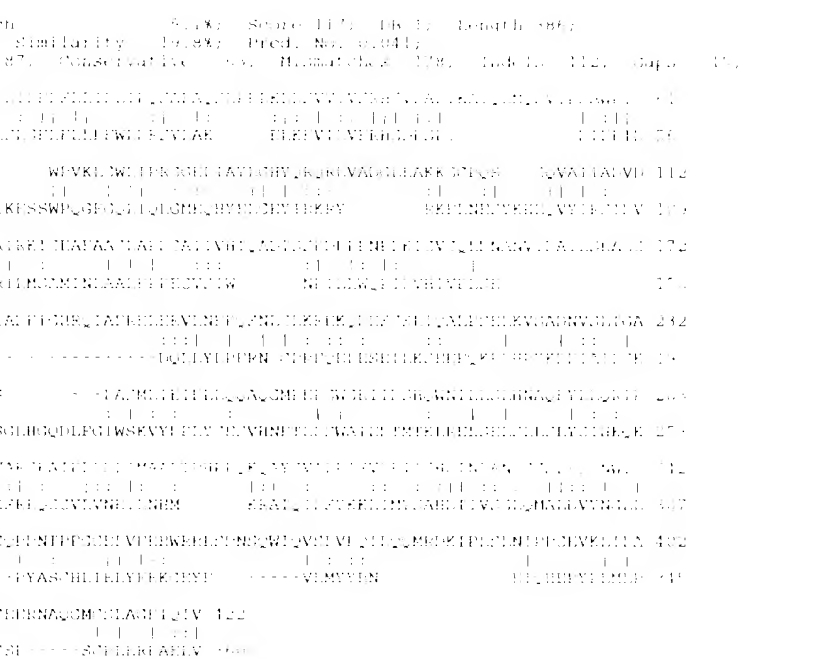
```
ID   "Human lysosomal acid phosphatase: cloning, expression and
AC   "Structural gene".";
DT   EMBL / 27-JAN-1988 / 2643 / 2650(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE: Leukocyte;
RX   MEDLINE; 88067928; PubMed 2772754;
RA   Setzer C., Zoller H., Pohlmann R.;
RT   "Structure of the human lysosomal acid phosphatase gene.";
RL   Eur. J. Biochem. 168:611-616(1987).
CC   -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN
CC       ALCOHOL + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: LYSOSOMAL.
CC   -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15029; CAA33042.1; -.
DR   EMBL; X15025; CAA33042.1; -.
DR   EMBL; X15026; CAA33042.1; JOINED.
DR   EMBL; X15027; CAA33042.1; JOINED.
DR   EMBL; X15028; CAA33042.1; JOINED.
DR   EMBL; X15029; CAA33042.1; JOINED.
DR   EMBL; X15030; CAA33042.1; JOINED.
DR   EMBL; X15031; CAA33042.1; JOINED.
DR   EMBL; X15032; CAA33042.1; JOINED.
DR   EMBL; X15033; CAA33042.1; JOINED.
DR   EMBL; X15034; CAA33042.1; JOINED.
DR   EMBL; X15035; CAA33042.1; JOINED.
DR   PIR; S00167; S00167.
DR   HSSP; P20645; 1RFE.
DR   MIM; 17599; -.
DR   InterPro; IPR000560; -.
DR   Pfam; PF00328; acid_phosphat_1.
DR   PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR   PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW   Hydrolase; Sialic; Glycoprotein; Lysosome.
FT   SIGNAL       1     40
FT   CHAIN       41    423       LYSOSOMAL ACID PHOSPHATASE.
FT   DISULFID   159    370       BY SIMILARITY.
FT   DISULFID   215    310       BY SIMILARITY.
FT   DISULFID   345    349       BY SIMILARITY.
FT   ACT_SITE    42     42       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    84     84       BY SIMILARITY
FT   ACT_SITE   289    289       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    92     92       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   133    133       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   167    167       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   177    177       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   191    191       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   257    257       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   322    322       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   411    411       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   423 AA;  47443 MW;  4F5A4A8B0A1E1B61 CRC64;

  Query Match             %;  Score 1x.0;  DB 1;  Length 423;
  Best Local Similarity   %;  Pred. No. 1.0e-%;
  Matches   99;  Conservative   41;  Mismatches  100;  Indels   19;  Gaps    1%;

Qy      1 MGVVVLVLFLWTAVEGLGMDVKLVESGGGLVQPGGSRKLSCAASGFTFSSFGMHWVRQAPEKGLEWVAY 70
          |:|  :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      33 LRFVSLLYLTESISGVKS--YEKEFPYIPEEWQMERIVQNLTEDHPMLMTVWSGEEPGSLERVCTPVYD 100

Qy     71 AISSGGSYTYYPDTVKGRFTISRDNAKNTLYLQMSSLRSEDTALYYCARQGYGNYVEAMDYWGQGTSVTV 140
          |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     101 AVSSGGSYTYYPDTVKGRFTISRDNAQSTLFLQMNSLRSEDTAMYYCVRHGNFGNSYISYWAYWGQGTLVTV 170

Qy    141 SAAKTTPPSVYPLAPGSAAQTNSMVTLGCLVKGYFPEPVTVTWNSGSLSSGVHTFPAVLQSDLYTLSSSV 210
          |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     171 SAKTTAPSVYPLAPVCGDTTGSSVTLGCLVKGYFPEPVTLTWNSGSLSSGVHTFPAVLQSDLYTLSSSV 240

Qy    211 TVPSSTWPSETVTCNVAHPASSTKVDKKIEPRGPTIKPCPPCKCPAPNLLGGPSVFIFPPKIKDVLMISL 280
          |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     241 TVPSSTWPSQTVTCSVAHPASSTTVDKKLEPSGPISTINPCPPCKECHKCPAPNLEGGPSVFIFPPNIKDVLMIAL 310

Qy    281 SPIVTCVVVDVSEDDPDVQISWFVNNVEVHTAQTQTHREDYNSTLRVVSALPIQHQDWMSGKEFKCKVNN 350
          |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     311 TPKVTCVVVDVSKDDPEVQFSWFVDDVEVHTAETKPREEQFNSTFRSVSELPIMHQDWLNGKEFKCRVNS 380

Qy    351 AALPAPIEKTISKTKGRPKAPQVYTIPPPKEQMAKDKVSLTCMITDFFPEDITVEWQWNGQPAENYKNTQ 420
          |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     381 AAFPAPIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTP 450

Qy    421 PVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSLGK 474
          |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     451 PVLDSDGSFFLYSKLTVDKSRWQQGNIFSCSVMHEALHNHYTQKSLSLSPGK 504

RESULT   6
PPAL_RAT
ID   PPAL_RAT        STANDARD;       PRT;   423 AA.
AC   P20611;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   LYSOSOMAL ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2) (LAP).
GN   ACP2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muroidea; Muridae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89250916; PubMed 2764916;
RA   Hibino M., Fujita H., Noguchi Y., Koko A., Kato K.;
RT   "Isolation and sequencing of a cDNA clone encoding acid phosphatase
RT   in rat liver lysosomes.";
RL   Biochem. Biophys. Res. Commun. 162:1044-1053(1989).
CC   -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN
CC       ALCOHOL + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: LYSOSOMAL.
CC   -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27049; AAA40744.1; -.
DR   PIR; A33995; A33995.
DR   HSSP; P20645; 1RFE.
DR   InterPro; IPR000560; -.
DR   Pfam; PF00328; acid_phosphat_1.
DR   PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR   PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW   Hydrolase; Signal; Glycoprotein; Lysosome.
FT   SIGNAL       1     40
FT   CHAIN       41    423       LYSOSOMAL ACID PHOSPHATASE.
FT   DISULFID   159    370       BY SIMILARITY.
FT   DISULFID   215    310       BY SIMILARITY.
FT   DISULFID   345    349       BY SIMILARITY.
FT   ACT_SITE    42     42       NUCLEOPHILE (BY SIMILARITY).
```

```
FT   ACT_SITE     84     84       BY SIMILARITY.
FT   ACT_SITE    28[illegible]    28[illegible]       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    [illegible]2    [illegible]2       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1[illegible]   1[illegible]       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   167    167       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   177    177       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   191    191       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   197    197       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   267    267       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   322    322       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   331    331       N LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   42[illegible] AA;  48[illegible] MW;  [illegible]F4C481[illegible]CAD[illegible]4196 CRC64;

  Query Match             5.7%;  Score 131;  DB 1;  Length 423;
  Best Local Similarity   22.8%;  Pred. No. 0.0035;
  Matches  1[illegible]5;  Conservative   45;  Mismatches  184;  Indels  122;  Gaps   23;

Qy          2 KAILIPFL [illegible] VVIV[illegible] 59
Db         1[illegible] QAAL[illegible] 58

Qy         60 PIWPVF[illegible] 118
Db         59 [illegible] 113

Qy        11[illegible] [illegible] 178
Db        114 AEANLAGLPF[illegible] FVHIVF[illegible] 14[illegible]

Qy        179 [illegible] 2[illegible]
Db        14[illegible] -------[illegible] 1[illegible]

Qy        2[illegible]9 [illegible] YLLQKITEVA 286
Db        195 LMNLTLH[illegible]IWNVY[illegible] 251

Qy        287 R[illegible]R-------A[illegible]MAAL[illegible]QKQAY[illegible]CALE[illegible]N 339
Db        252 KARL[illegible]VLLA[illegible]KN[illegible]MAL[illegible]PPK ---  ----  LIVYCAHD[illegible]VAL[illegible]MALNVY 300

Qy        34[illegible] W[illegible]L[illegible]LVPERWR[illegible]EVKL 399
Db        301 -----NGK[illegible]EAPW ----  --PL 340

Qy        4[illegible]2 [illegible]ERNA[illegible]VNEARIP  ACLLRS 434
Db        341 [illegible]PGCPHR---- [illegible]KEQLAS 374

RESULT  7
PPAL_MOUSE
ID   PPAL_MOUSE    STANDARD;    PRT;   42[illegible] AA.
AC   P24638;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   LYSOSOMAL ACID PHOSPHATASE PRECURSOR (EC [illegible]) (LAP) (FRAGMENT).
GN   ACP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91282986; PubMed=2059337;
RA   [illegible] Pfl[illegible] E., [illegible]
RT   "Molecular cloning of the mouse lysosomal acid phosphatase.";
RL   Biol. Chem. Hoppe Seyler 372:301-304(1991).
CC   -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN
CC       ALCOHOL + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: LYSOSOMAL.
CC   -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X[illegible]; CAA40485.1; -.
DR   PIR; S14742; S14742.
DR   HSSP; P20646; 1RPT.
DR   MGD; MGI:87882; Acp2.
DR   InterPro; IPR000560; -.
DR   Pfam; PF00328; acid_phosphat; 1.
DR   PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR   PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW   Hydrolase; Signal; Glycoprotein; Lysosome.
FT   NON_TER       1      1
FT   SIGNAL        1     28
FT   CHAIN        29    421       LYSOSOMAL ACID PHOSPHATASE.
FT   DISULFID    [illegible]    [illegible]       BY SIMILARITY.
FT   DISULFID    210    308       BY SIMILARITY.
FT   DISULFID    [illegible]    417       BY SIMILARITY.
FT   ACT_SITE     40     40       NUCLEOPHILIC ACCEPTOR (BY SIMILARITY)
FT   ACT_SITE    [illegible]    [illegible]       BY SIMILARITY.
FT   ACT_SITE    284    284       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD     [illegible]     [illegible]       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1[illegible]    1[illegible]       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1[illegible]    1[illegible]       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    175    175       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1[illegible]    1[illegible]       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    195    195       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    267    267       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    320    320       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    3[illegible]    3[illegible]       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    372    372       N LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   421 AA;  48279 MW;  C66[illegible]F431239D676D CRC64;

  Query Match             5.3%;  Score 122.5;  DB 1;  Length 421;
  Best Local Similarity   21.8%;  Pred. No. 0.017;
  Matches   97;  Conservative   49;  Mismatches  191;  Indels  10[illegible];  Gaps   19;

Qy          2 KAILIPF  LSFLLPFTFQ[illegible]AFAQSEPELKLESVVIVSRHGVRAPTKAT[illegible]M[illegible]V[illegible] 5[illegible]
Db          8 QAALLQFLLGMCLTVMPPIQARS        LRFVTLLYRHGDRSPVKT    YPKDP 53

Qy         5[illegible] DAN[illegible]VP[illegible]VA[illegible]LAKK[illegible]VA[illegible]V[illegible] 11[illegible]
Db         54 YQEEKW[illegible]QL[illegible]KEN[illegible]QAL[illegible]RY  HGFL   NTSYHRQEVYVRSTDFDRT 108

Qy        11[illegible] [illegible]A 17[illegible]
Db        109 LMSAEANLAGLFPP   NEV[illegible]HESPN[illegible]WQPI    PVHIVPEIE    1[illegible]

Qy        17[illegible] [illegible] 2[illegible]
Db        147             DRLLKFPLG[illegible]PRYE[illegible]LQND    FRQTPEYQNR[illegible]LNA[illegible]NNVAN 18[illegible]

Qy        236 ASHL[illegible]   L[illegible]QA[illegible]QMPEP[illegible]WGKI[illegible]DSHQWNILLSLHNA[illegible]FY[illegible] 284
Db        190 EPGLTNVILLETIWNVYPELLFCRQTHGLLLPPWASPQIVQRLSQLKDFSPLPLLCIHR[illegible]VQ 249

Qy        2[illegible] [illegible]NWI[illegible] 3[illegible]
Db        2[illegible]0 [illegible]NVAL[illegible]MALNVY   [illegible]

Qy        3[illegible]5 Q[illegible]NTP[illegible]LVPERWRRL[illegible]WIQVC[illegible]LVP[illegible]MRPKIP[illegible]LNIPPGEVKLT[illegible]LAC[illegible] 404
Db        2[illegible]9 N[illegible]K[illegible]E[illegible]NF  [illegible]N[illegible]  EAPWLLLP    [illegible]
```

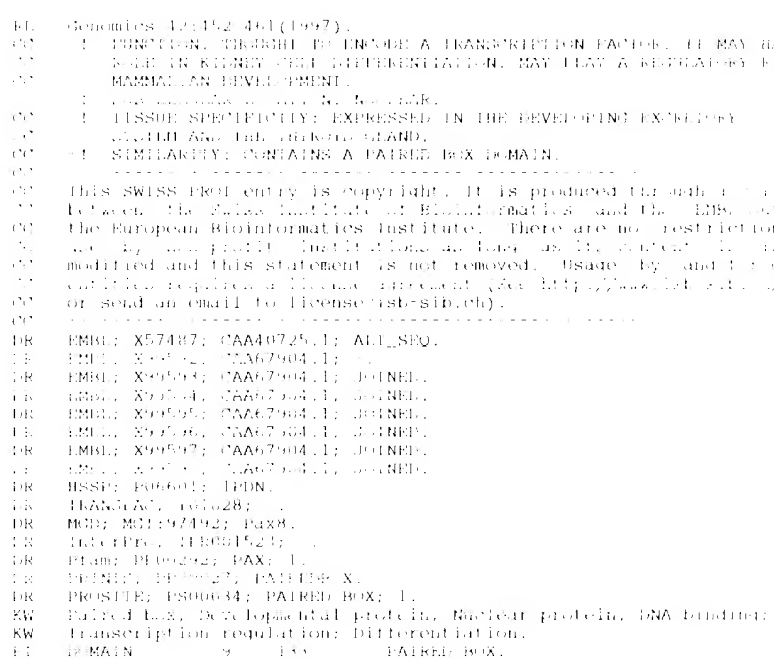
```
XX   10% RRNA, 7% SRNA AND 83% UNSPECIFIC DNA
XX   ----------|----|----|----|----|----|----|----|----|----|
DE   43 FBP      CHLOROPLAST RIBOSOMAL
```

```
RESULT  8
PPAP_HUMAN
ID   PPAP_HUMAN      STANDARD;      PRT;   402 AA.
AC   P13686;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-MAY-1990 (Rel. 15, Last sequence update)
DT   01-OCT-2000 (Rel. 39, Last annotation update)
DE   PROSTATIC ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2).
GN   ACPP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92208413; PubMed=1371434;
RA   Sharief F.S., Li S.S.-L.;
RT   "Structure of human prostatic acid phosphatase gene.";
RL   Biochem. Biophys. Res. Commun. 184:1158-1170(1992).
RN   [2]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.
RX   MEDLINE=91119038; PubMed=1987808;
RA   van Etten R.L., Davidson R., Stevis P.E., Macarthur H., Moore D.L.;
RT   "Covalent structure, disulfide bonding, and identification of
RT   reactive surface and active site residues of human prostatic acid
RT   phosphatase.";
RL   J. Biol. Chem. 266:2313-2319(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90262784; PubMed=2171434;
RA   Sharief F.S., Lee H., Deaderman M.M., Dundwal I.A., Lee C.L.,
RA   Lee C., Li S.S.-L.;
RT   "Human prostatic acid phosphatase: cDNA cloning, gene mapping and
RT   protein sequence homology with lysosomal acid phosphatase.";
RL   Biochem. Biophys. Res. Commun. 169:79-86(1990).
RN   [4]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Prostate;
RX   MEDLINE=88312981; PubMed=2841284;
RA   Vihko P., Virkkunen P., Henttu P., Roiko K., Solin T., Huhtala M.L.,
RT   "Molecular cloning and sequence analysis of cDNA for human
RT   prostatic acid phosphatase.";
RL   FEBS Lett. 236:275-281(1988).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=90172161; PubMed=2405574;
RA   Lebel J.L., Goodrum M.V., Katzel P.C.,
RT   "Nucleotide sequence of human prostatic acid phosphatase determined
RT   from a full length cDNA clone.";
RL   Nucleic Acids Res. 18:1054-1054(1990).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95108554; PubMed=7961974;
RA   Sharief F.S., Li S.S.-L.;
RT   "Nucleotide sequence of human prostatic acid phosphatase ACPP gene.
RT   Including a new Alu repeat.";
RL   Biochem. Mol. Biol. Int. 35:951-960(1994).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE=99139966; PubMed=9918385;
RA   Lacour M.W., Banas T., Ichtchenko I.,
RT   "2.0 Angstrom structure of L(+) tartrate inhibited human prostatic
RT   acid phosphatase.";
RL   J. Biol. Chem. 274:3004-3009(1999).
CC   -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN
CC       ALCOHOL + ORTHOPHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER.
```

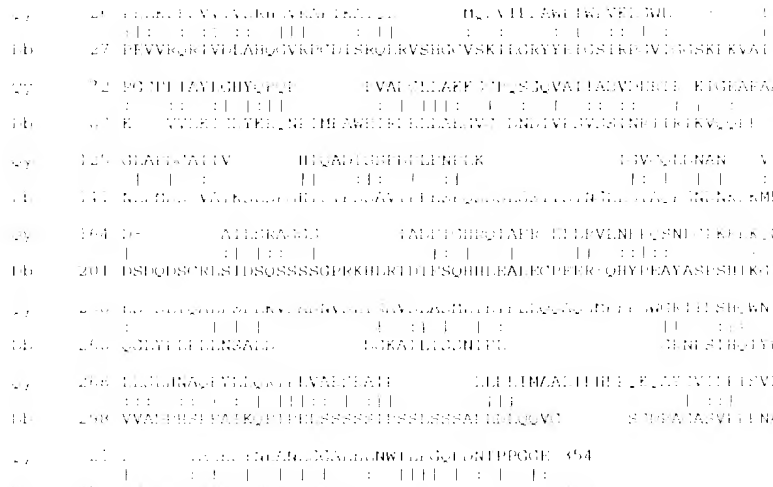
```
CC   -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27569; AAA60021.1; -.
DR   EMBL; M97580; AAA60021.1; JOINED.
DR   EMBL; M97581; AAA60021.1; JOINED.
DR   EMBL; M97582; AAA60021.1; JOINED.
DR   EMBL; M97583; AAA60021.1; JOINED.
DR   EMBL; M97584; AAA60021.1; JOINED.
DR   EMBL; M97585; AAA60021.1; JOINED.
DR   EMBL; M97586; AAA60021.1; JOINED.
DR   EMBL; M97587; AAA60021.1; JOINED.
DR   EMBL; M97588; AAA60021.1; JOINED.
DR   EMBL; M34401; AAA60664.1; -.
DR   EMBL; M34902; AAA60022.1; -.
DR   EMBL; X52174; CAA36422.1; -.
DR   EMBL; X56605; CAA39674.1; -.
DR   EMBL; D07997; AAB60640.1; -.
DR   EMBL; D07083; AAB60640.1; JOINED.
DR   EMBL; D07085; AAB60640.1; JOINED.
DR   EMBL; D07086; AAB60640.1; JOINED.
DR   EMBL; D07088; AAB60640.1; JOINED.
DR   EMBL; D07091; AAB60640.1; JOINED.
DR   EMBL; D07092; AAB60640.1; JOINED.
DR   EMBL; D07093; AAB60640.1; JOINED.
DR   EMBL; D07095; AAB60640.1; JOINED.
DR   PIR; A32419; A32419.
DR   PIR; S01331; S01331.
DR   PIR; S11147; S11147.
DR   PIR; JH0610; JH0610.
DR   PDB; 2HPA; 10-SEP-98.
DR   MIM; 171790; -.
DR   InterPro; IPR000560; -.
DR   Pfam; PF00328; acid_phosphat; 1.
DR   ProSite; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR   ProSite; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW   Hydrolase; Glycoprotein; Signal; 3D-structure.
FT   SIGNAL       1     32
FT   CHAIN       33    386       PROSTATIC ACID PHOSPHATASE.
FT   DISULFID    141    372
FT   DISULFID    215    313
FT   DISULFID    347    351
FT   ACT_SITE     44     44       NUCLEOPHILIC ATTACK (BY SIMILARITY).
FT   ACT_SITE     89     89       BY SIMILARITY.
FT   ACT_SITE    259    259       PROTON DONOR.
FT   CARBOHYD     94     94       N-LINKED (GLCNAC...).
FT   CARBOHYD    220    220       N-LINKED (GLCNAC...).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC...).
FT   CONFLICT     13     24       SLGFLFLLFF -> AFASCFCFFC (IN REF. 4
FT                                AND 6).
FT   CONFLICT     15     24       SGFLFLLFF -> AFASCFCFFC (IN REF. 5).
FT   CONFLICT     42     46       D -> H (IN REF. 5).
FT   CONFLICT     66     73       GPGQLTQL -> WTWPTHPA (IN REF. 1).
FT   CONFLICT     66     73       GPGQLTQL -> WTWPTHPA (IN REF. 1).
FT   CONFLICT     95     95       E -> D (IN REF. 4).
FT   CONFLICT    115    115       A -> R (IN REF. 4).
FT   CONFLICT    139    139       Q -> E (IN REF. 5).
FT   CONFLICT    157    157       F -> R (IN REF. 5).
FT   CONFLICT    212    212       P -> A (IN REF. 4).
FT   CONFLICT    215    215       C -> S (IN REF. 5).
FT   CONFLICT    294    294       S -> I (IN REF. 5).
FT   CONFLICT    342    342       C -> V (IN REF. 5).
FT   CONFLICT    383    383       D -> N (IN REF. 5).
SQ   SEQUENCE   386 AA;  44566 MW;  EFB1EFFFA3C9ADEA CRC64;
```

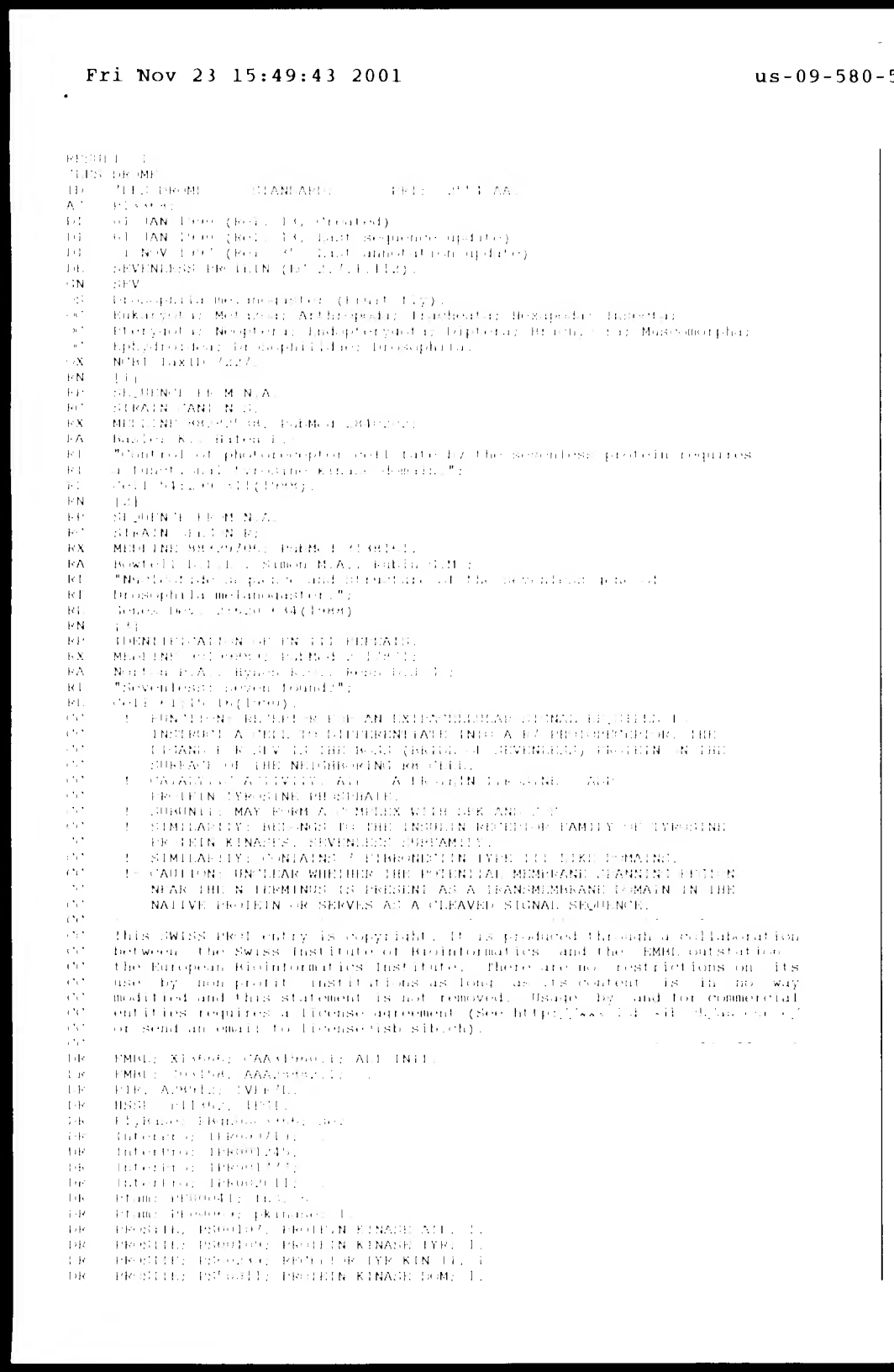


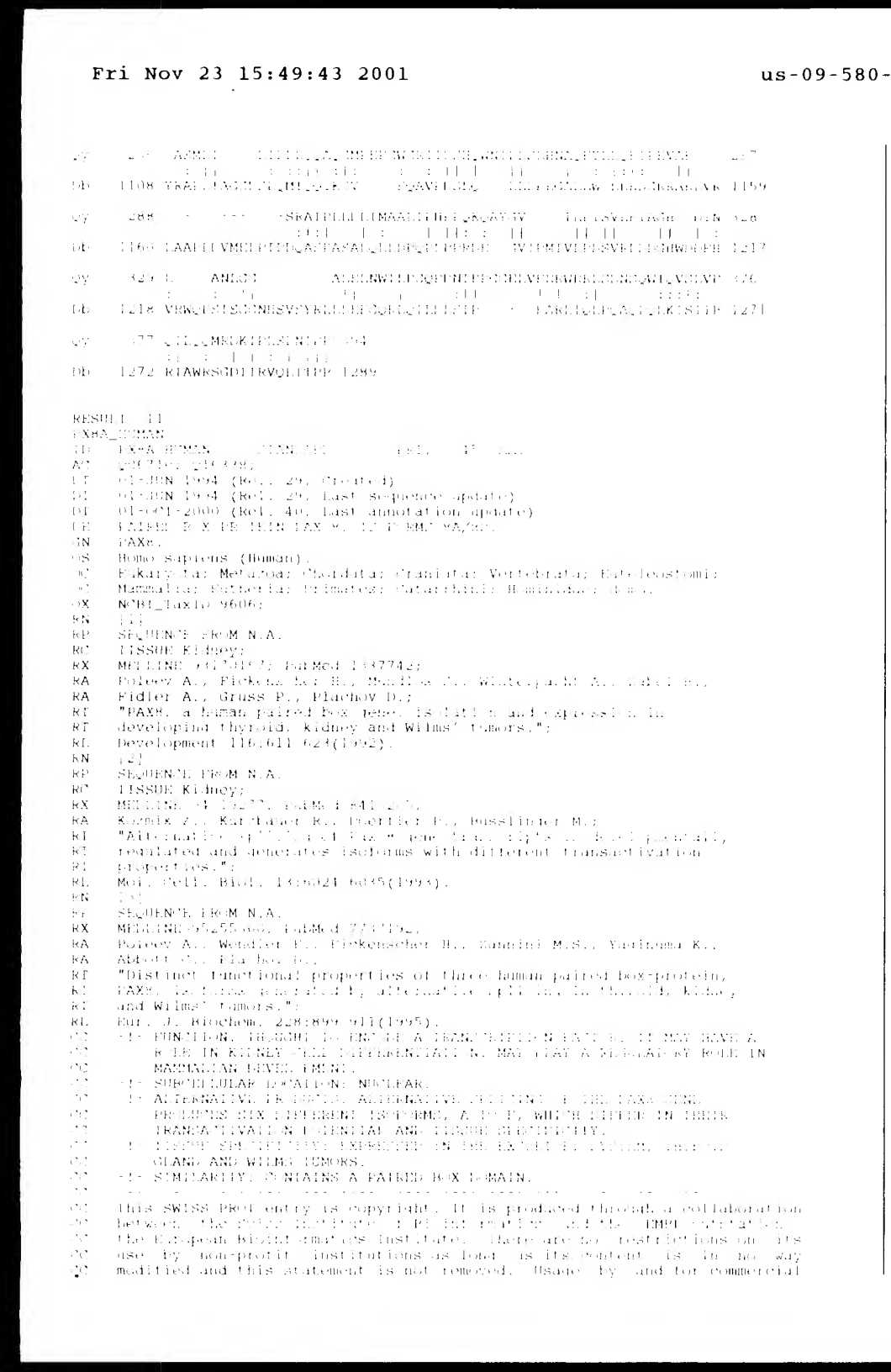
RESULT 9
PAX8_MOUSE
ID PAX8_MOUSE STANDARD; PRT; 457 AA.
AC Q00288; P97342;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PAIRED BOX PROTEIN PAX-8.
GN PAX8 OR PAX-8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92155118; PubMed=1730650;
RA Plachov D., Chowdhury K., Walther C., Simon D., Guenet J.-L.,
RA Gruss P.;
RT "PAX8, a murine paired box gene expressed in the developing kidney.";
RL Development 110:643-651(1990).
RN [2]
RP REVISIONS TO 158, 261 AND 340.
RX MEDLINE=94268737; PubMed=7913925;
RA Adams B., Dorfler P., Aguzzi A., Kozmik Z., Urbanek P.,
RA Maurer-Fogy I., Busslinger M.;
RT "Pax 5 encodes the transcription factor BSAP and is expressed in B
RT lymphocytes, the developing CNS, and adult testis.";
RL Genes Dev. 6:1589-1607(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Eye;
RX MEDLINE=97349113; PubMed=9205117;
RA Kikkawa Y., Fukuda A., Tamura S., Ito M., Hirabayashi Y., Hirai H.;
RT "Structural organization of the murine Pax 8 gene and
RT characterization of its basal promoter.";



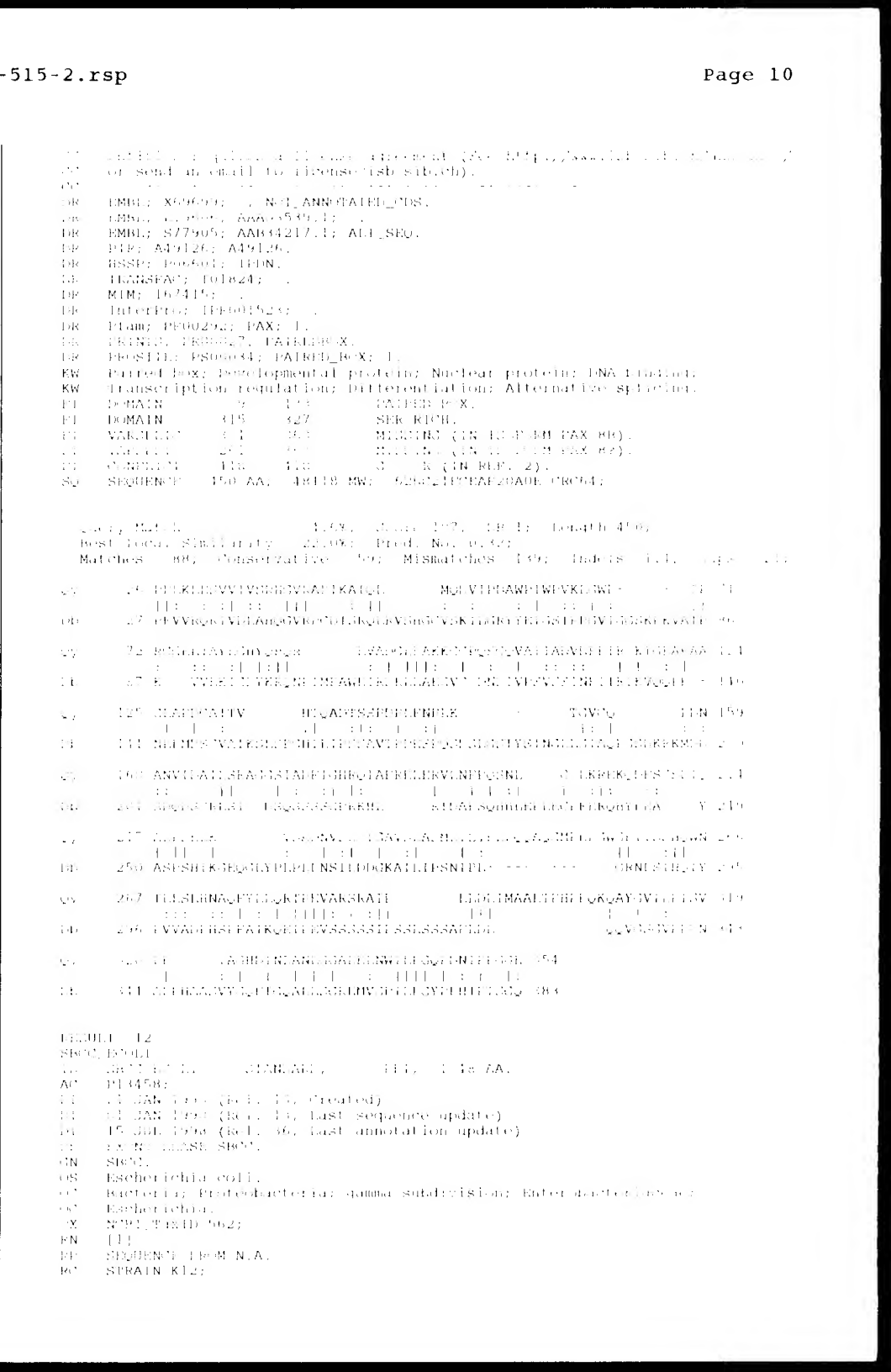
RL Genomics 12:152-161(1997).
CC -!- FUNCTION: THOUGHT TO ENCODE A TRANSCRIPTION FACTOR. IT MAY HAVE A
CC SPECIFIC ROLE DURING CELL DIFFERENTIATION, MAY PLAY A REGULATORY ROLE IN
CC MAMMALIAN DEVELOPMENT.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE DEVELOPING EXCRETORY
CC SYSTEM AND THYROID GLAND.
CC -!- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; X57487; CAA40725.1; ALL_SEQ.
DR EMBL; X99593; CAA67904.1; JOINED.
DR EMBL; X99594; CAA67904.1; JOINED.
DR EMBL; X99595; CAA67904.1; JOINED.
DR EMBL; X99596; CAA67904.1; JOINED.
DR EMBL; X99597; CAA67904.1; JOINED.
DR MGD; MGI:97492; Pax8.
DR InterPro; IPR001523; .
DR Pfam; PF00292; PAX; 1.
DR SMART; SM00351; PAIRED; 1.
DR PROSITE; PS00034; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA binding;
KW Transcription regulation; Differentiation.
FT DOMAIN 9 133 PAIRED BOX.
FT DOMAIN 316 328 SER-RICH.
FT CONFLICT 340 349 D -> G (IN REF. 3).
SQ SEQUENCE 457 AA; 48651 MW; 0AAC3ED54AB60AE CRC64;







RESULT 11
PAX6_HUMAN
ID PAX6_HUMAN STANDARD; PRT; 422 AA.
AC P26367; Q16439;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PAIRED BOX PROTEIN PAX-6. (AN)
GN PAX6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=91270157; PubMed=1847424;
RA Poliev A., Fickenscher H., Mueller S., Wolf E., Gruss P., Dahl E.,
RA Fidler A., Gruss P., Plachov D.;
RT "PAX6, a human paired box gene: isolation and expression in
RT developing thyroid, kidney and Wilms' tumors.";
RL Development 116:611-623(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94[illegible];
RA Kozmik Z., Kurzbauer R., Dorfler P., Busslinger M.;
RT "Alternative [illegible] in the human paired box (Pax) gene
RT regulated and generates isoforms with different transactivation
RT properties.";
RL Mol. Cell. Biol. 13:6021-6035(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95[illegible];
RA Polimeno A., Weadick P., Fickenscher H., Bamini M.S., Yoshioka K.,
RA Abbott C., Plachov D.;
RT "Distinct functional properties of three human paired box-protein,
RT PAX6, isoforms generated by alternative splicing in thyroid, kidney,
RT and Wilms' tumors.";
RL Eur. J. Biochem. 228:899-911(1995).
CC -!- FUNCTION: THOUGH IT IS EXPRESSED AT [illegible] IT MAY HAVE A
CC ROLE IN KIDNEY AND DIFFERENTIATION. MAY PLAY A DEVELOPMENTAL ROLE IN
CC MAMMALIAN INNER EAR.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE SPLICING: ALTERNATIVE SPLICING [illegible] THE PAX6 GENE
CC PRODUCTS SIX DIFFERENT ISOFORMS, A TO F, WHICH DIFFER IN THEIR
CC TRANSACTIVATION POTENTIAL AND DNA-BINDING SPECIFICITY.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE EARLY [illegible] EYE,
CC OLFACTORY AND WILMS TUMORS.
CC -!- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
CC -!- [illegible]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the [illegible] and the EMBL outstation - the
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial



CC or send an email to license@isb-sib.ch).
CC -!- [illegible]
DR EMBL; X69999; [illegible]; NOT_ANNOTATED_CDS.
DR EMBL; [illegible]; AAA[illegible]; -.
DR EMBL; S77905; AAB34217.1; ALT_SEQ.
DR PIR; A49126; A49126.
DR HSSP; P02465; 1HDN.
DR HGNREAC; [illegible]; -.
DR MIM; 167415; -.
DR InterPro; IPR001523; -.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PR00027; PAIRED_BOX; 1.
DR PROSITE; PS00034; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA binding;
KW Transcription regulation; Differentiation; Alternative splicing.
FT DOMAIN 9 128 PAIRED BOX.
FT DOMAIN 415 427 SER RICH.
FT VARSPLIC 1 84 MISSING (IN ISOFORM PAX-8B).
FT [illegible] 85 [illegible] MISSING (IN ISOFORM PAX-8A).
FT CONFLICT 116 118 S -> R (IN REF. 2).
SQ SEQUENCE 150 AA; 18110 MW; 02A6C_1D7EAF20ADE CRC64;

[illegible] Matches 139; Conservative 0; Mismatches 139; Indels 1; Gaps 1;
Best Local Similarity 22.0%; Pred. No. 0.42;
Matches 88; Conservative 30; Mismatches 139; Indels 11; Gaps 11;

Qy 24 PEIKLCVVIVSGHCVSAEIKAFCKS MQEVIPSAWPFIWFVKGLWT 73
[illegible]
Db 27 PEVVRQAGYKLGAMQVVLGGLRVAPRTLVKPEVVAAQIVQYLGGRQTSVTL 76

Qy 74 PGGELQAAQHHYAYEGA SAVLTGSPAPAVSFEVARGDEPIPKTGVAAA 123
[illegible]
Db 77 E [illegible] PEEIKQRIVELHKQENPTWAQDGSLRETCGLSKINNRIIETWYSDNLKAMV 126

Qy 124 QSAPPVAHTV HEVASPSPPPLPNPLE Y TGVDV IIN 159
[illegible]
Db 127 SAIMESVVAHGSLTPCDISVIVNGYLFSNPCLSIRKSRMGGVCFMVADDWGV 176

Qy 160 ANVTLAILSFACFSSIALPTSHRQLIAFNELPEVLNFPVSNL S LKPEKLIFS 211
[illegible]
Db 177 SRVSSLLKPFLKS LSQDSGSGGEEHSL SSIAFSGMNNLTDTFTSQQITA Y 219

Qy 212 [illegible] SASVSL [illegible] 259
[illegible]
Db 220 ASPSHIKSPLQLGLPLPLINSIIDGKAILIPSNIPL--- --- GKNISIHLSY 259

Qy 260 TLSSLHNAQFTLSKTEVAKSKAIT LLDGIMAALFPHIQKQAYSVTLTLV 310
[illegible]
Db 260 IVVADHRSLFAIKQEIPEVESSSSSIFSSLSSSAFLDI QLVSSNVIP N 310

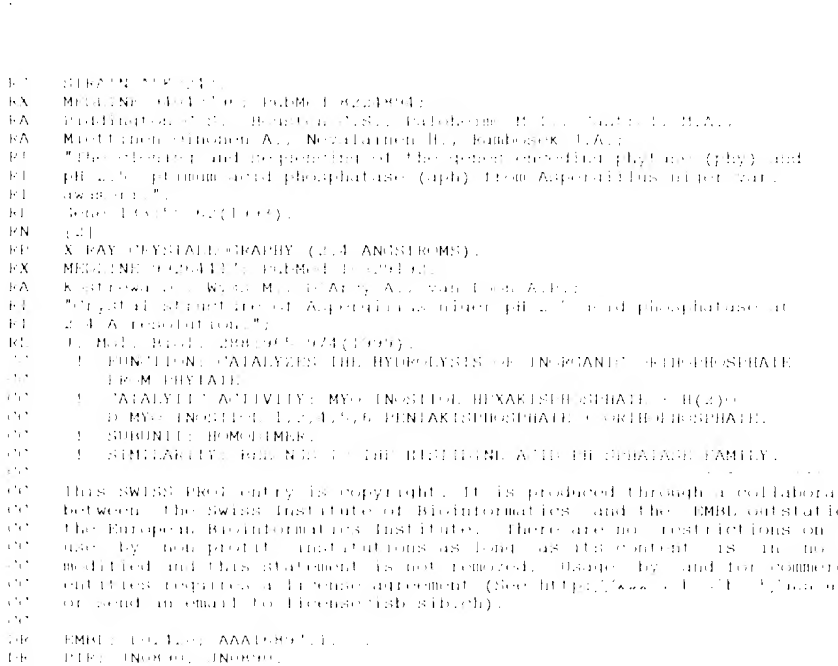
Qy [illegible] 311
[illegible]
Db 311 SEIHGAVVSLFLQLALSAGLNWSPSLGYTFHIFGMQP 344

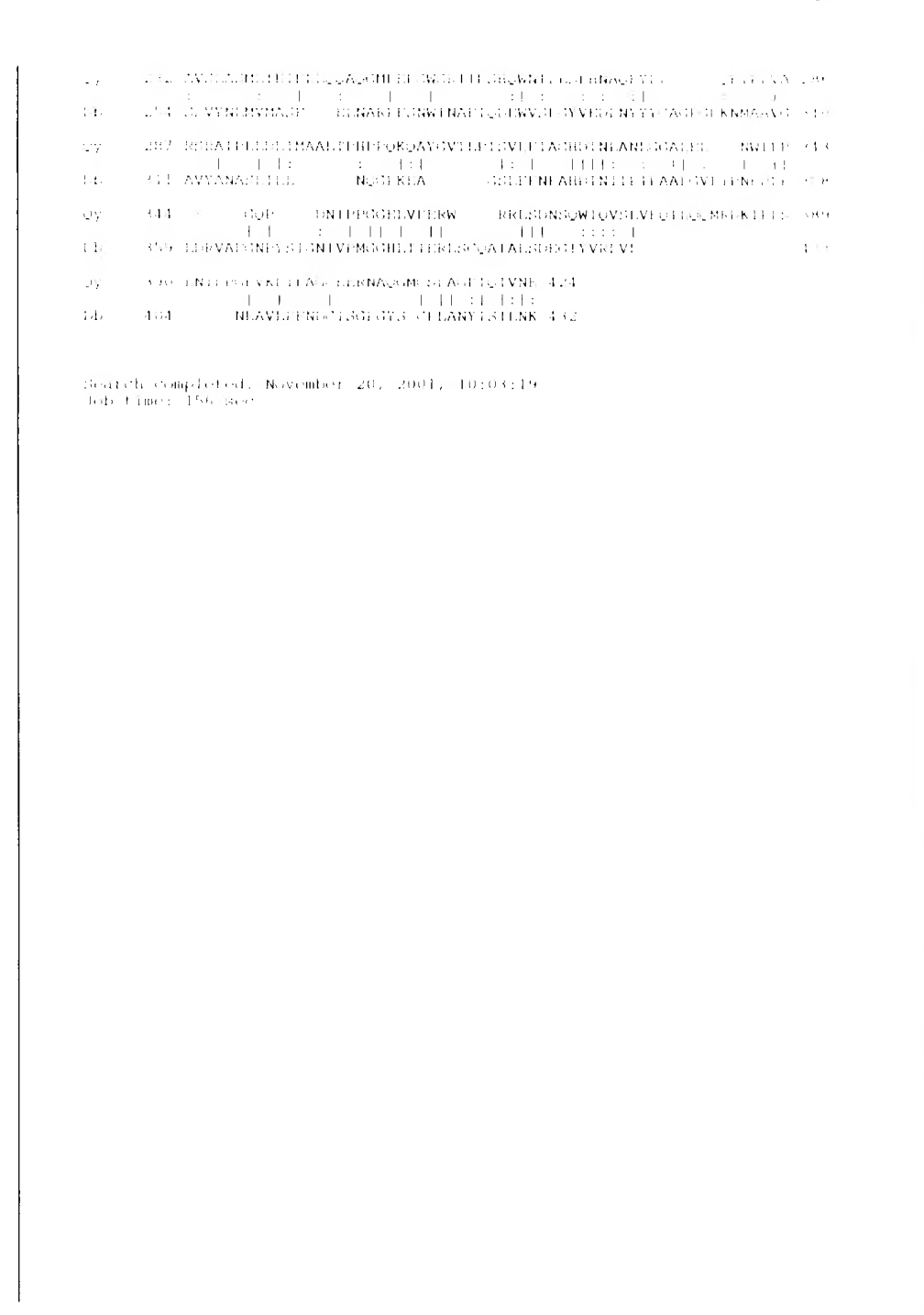
RESULT 12
SB50_ECOLI
[illegible] STANDARD; PRT; [illegible] AA.
AC P18458;
DT [illegible] (Rel. 15, Created)
DT [illegible] (Rel. 15, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE [illegible] SB50.
GN SB50.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gamma subdivision; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;

```
RX   MEDLINE; 96015041; PubMed=8557671;
RA   Norris V., Grant S., Mottes M., Root C., Beach D.H.F., Lloyd R.G.;
RT   "Molecular organization of sbcC, a gene that affects genetic
RT   recombination and the stability of DNA palindromes in Escherichia
RT   coli K-12.";
RL   Nucleic Acids Res. 23:4553-4560(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94112001; PubMed=8282603;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Ivanova N., Aleshin V., Artamonova I., Apartsina A.M., Tutukina M.,
RA   Fedorkpsted N., Hyman R., Kaiman S., Kemp C., Kurdi O., Law H.,
RA   Lin D., Namath A., Oechner P., Roberts D., Schramm S., Tarvis R.W.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96044966; PubMed=7744255;
RA   Reeder T., Schleif R.;
RT   "Mapping, sequence, and apparent lack of function of araI, a gene of
RT   the Escherichia coli ara operon.";
RL   J. Bacteriol. 177:5705-5711(1995).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=94134410; PubMed=8106031;
RA   Leach D.R.F., Lloyd R.G., Connison A.F.;
RT   "The SbcCD protein of Escherichia coli is a, fated to two-jointed
RT   nucleases in the MutA superfamily of nucleotide binding proteins.";
RL   Genetica 87:65-100(1992).
CC   -!- FUNCTION: AFFECTS GENETIC RECOMBINATION AND THE VIABILITY OF
CC       DNA PALINDROMES IN E.COLI. SBCC MAY HAVE A NUCLEASE ACTIVITY THAT
CC       IS FUNCTIONALLY RELATED TO ONE OF THE NUCLEASE ACTIVITIES OF THE
CC       RECBCD ENZYME.
CC   -!- SIMILARITY: TO PHAGE P4 GENE 46 PROTEIN AND PHAGE 15 GP1 AND TO
CC       UVRA/RECN.
CC
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15001; CAA33104.1;
DR   EMBL; AE000111; AAC73001.1;
DR   EMBL; U14003; AAB96037.1;
DR   EMBL; M74017; AAA24177.1;
DR   PIR; I36056; RVECSC;
DR   PIR; S27548; S27548.
DR   PROSITE; PS00017; ATP_GTP_A;
KW   Hydrolase; Nuclease; ATP-binding; Coiled coil;
FT   NP-BIND      37     44       ATP (POTENTIAL).
FT   DOMAIN       191    290      COILED COIL (POTENTIAL).
FT   DOMAIN       307    402      COILED COIL (POTENTIAL).
FT   DOMAIN       438    502      COILED COIL (POTENTIAL).
FT   DOMAIN       528    596      COILED COIL (POTENTIAL).
FT   DOMAIN       607    684      COILED COIL (POTENTIAL).
FT   DOMAIN       721    797      COILED COIL (POTENTIAL).
FT   DOMAIN       820    881      COILED COIL (POTENTIAL).
FT   DOMAIN       900    980      COILED COIL (POTENTIAL).
SQ   SEQUENCE   1048 AA;  118721 MW;  BDB7AB7C3B0EF0C4 CRC64;

  Query Match          100.0%;  Score 1048;  DB 1;  Length 1048;
  Best Local Similarity   100.0%;  Pred. No. 0.0;
```

```
  Matches 1048;  Conservative   54;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      19 SAPAQSEPSSLKESVVEVSRRCVPAP  IEAIQL  MQDVIELAWPTWRVK  WIIRC 78
           :| ||    :  |    :|: ||   |   :  ||   :|:||:              :
Db     117 SALAGRALSSLSL RLNVALGSLIVPQQKRLAGLGVAIGNVIQPQ              LGN 166

Qy      79 GPEIAYSLHPQGKQKLVAIKSLLAKSGTLQSGQVAELAGVTRPTRKTGRAFAA  Ag  138
           |    : |||         |:   |  | |  : : |  || || |  |       |
Db     167 AAI    NGMRGFY      EEKEPGLIAGVEIEIGGARI RTLGAVPAQLGAS LGIG 216

Qy     139 IAVLVTQAESSGPHSELNPELKIGVPQ      LGNANVIGAIIRACISIAGL PNPL 198
           |  | |  :         : |: || |   ||         :  :  |  |  |
Db     217 LPKQIQKGENEAQS      LRQDE      QALIQQWQAVIASENII      LQPIVIQ 266

Qy     199 PWLLAQGEHEHQLRLLSQKHEEDQQIAAHNQQ  IDQYQQQIEQDPQQLLEIELGVAEE 258
           ||   |    :: |:    :: || |  ||   :  :  :  |  |  |   |
Db     267 PWILAQGEHEHQLRLLSQKHEEDQQIAAHNQQ  IDQYQQQIEQDPQQLLEIELGVAEE 326

Qy     259 EWRWRLSDNSQWIQVSLVFQTIQQ 281
           | |||:  : : :    |||||
Db     327 EWKWKVE GEQFALAGCQQQPLQQ 350

RESULT 13
PAX8_CANFA
ID   PAX8_CANFA     STANDARD;      PRT;   450 AA.
AC   P47240;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PAIRED BOX PROTEIN PAX-8, ISOFORM 8A.
GN   PAX8 OR PAX-8.
OS   Canis familiaris (dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Thyroid;
RX   MEDLINE=96354068; PubMed=8652674;
RA   [illegible], Vassart G., Christophe D.;
RT   "Pax-8 expression in primary cultured dog thyrocyte is increased by
RT   cyclic AMP.";
RL   Biochim. Biophys. Acta 1307:97-103(1996).
CC   -!- FUNCTION: [illegible] IS INVOLVED IN PAX8-A INTERACTION WITH PAX5 AND
CC       ROLE IN KIDNEY CELL DIFFERENTIATION. MAY PLAY A REGULAR BY E.G. IN
CC       MAMMALIAN DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: ALTERNATIVE SPLICING OF THE PAX8 GENE
CC       PRODUCES DIFFERENT ISOFORMS.
CC   -!- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X86761; CAA60711.1;
DR   HSSP; P06601; 1FHN.
DR   InterPro; IPR001523;
DR   Pfam; PF00292; PAX; 1.
DR   PRINTS; PR00027; PAIREDBOX.
```

```
ID   PR_SILI; [illegible]; [illegible].
KW   [illegible]; [illegible]; [illegible]; Nucleus; [illegible]; DNA-binding;
KW   Transcription regulation; Differentiation; Alternative splicing;
FT   DOMAIN          9     133       PAIRED DOMAIN.
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible] CRC64;


  Query Match             [illegible]%;  Score [illegible];  DB 1;  Length 159;
  Best Local Similarity   [illegible]%;  Pred. No. 0.44;
  Matches   [illegible];  Conservative   [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps   [illegible];

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

RESULT  14
SRPR_HUMAN
ID   SRPR_HUMAN      [illegible]      [illegible]
AC   [illegible];
DT   [illegible] (Rel. [illegible], Created)
DT   01-NOV-1998 (Rel. 08, Last sequence update)
DT   01-FEB-1996 (Rel. 34, Last annotation update)
DE   SIGNAL RECOGNITION PARTICLE RECEPTOR ALPHA SUBUNIT (SR ALPHA)
DE   (DOCKING PROTEIN ALPHA) (FT ALPHA).
GN   SRPR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88124220; PubMed=3340536;
RA   Hortsch M., Labeit S., Meyer D.I.;
RT   "Complete cDNA sequence coding for human docking protein.";
RL   Nucleic Acids Res. [illegible]:[illegible] (1988).
CC   -!- FUNCTION: [illegible] INTEGRAL MEMBRANE PROTEIN INSERTED [illegible] IN A POSTTRANSLATIONAL
CC       WITH SRP9. THE GTPASE ACTIVITY OF THE NASCENT SECRETED BY PROTEINS
CC       TO THE ENDOPLASMIC RETICULUM MEMBRANE SYSTEM.
CC   -!- SUBUNIT: BETA-SUBUNIT IS AN ALPHA AND A BETA CHAIN.
CC   -!- SUBCELLULAR LOCATION: [illegible]. [illegible] [illegible] [illegible] [illegible]
CC       BE ANCHORED IN THE MEMBRANE. THE [illegible] AN INTEGRAL [illegible]
CC       WHICH CONTAINS A [illegible] TRANSMEMBRANE [illegible].
CC   -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF SRP [illegible] [illegible].
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a particular agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X06439; CAA29608.1; -.
DR   PIR; A29440; A29440.
DR   HSSP; P07347; 1EFH.
DR   MIM; 182180; -.
DR   InterPro; IPR000897; -.
DR   Pfam; PF00448; SRP54; 1.
DR   PROSITE; PS00300; SRP54; 1.
KW   Signal recognition particle; Transmembrane; Receptor;
KW   Endoplasmic reticulum; GTP-binding;
FT   DOMAIN          11     34       POTENTIAL.
FT   NP_BIND      425    432       GTP (BY SIMILARITY).
FT   [illegible]      [illegible]    [illegible]  (BY SIMILARITY).
FT   NP_BIND      588    591       GTP (BY SIMILARITY).
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible] CRC64;


  Query Match             [illegible]%;  Score 102;  DB 1;  Length 648;
  Best Local Similarity   [illegible]%;  Pred. No. 1.1;
  Matches   [illegible];  Conservative   [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps   [illegible];

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

Qy      [illegible]

Db      [illegible]

RESULT  15
PHYA_ASPAW
ID   PHYA_ASPAW      [illegible]      [illegible]
AC   P34759;
DT   01-FEB-1995 (Rel. 28, Created)
DT   01-FEB-1996 (Rel. 34, Last sequence update)
DT   [illegible] (Rel. 39, Last annotation update)
DE   [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]
DE   [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]
GN   PHYA OR APH.
OS   Aspergillus awamori.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=105351;
RN   [1]
RP   SEQUENCE FROM N.A.
```





SW nucleic - nucleic search, using sw model

Run on: November 23, 2001, 12:17:37 ; Search Time 1.24 seconds
(without alignments)
5656.429 Million cell updates/sec

Title: US-09-580-515-1
Perfect score: 1423
Sequence: 1 ttttaaaaaatatttttttttt ... gatcaaaaaaaaaaaaaaaa 1423

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2381 seqs, 813650859 residues

Total number of hits satisfying chosen parameters: 1460204

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_200[1]*
1: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[1].seq:*
2: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[2].seq:*
3: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[3].seq:*
4: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[4].seq:*
5: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[5].seq:*
6: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[6].seq:*
7: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[7].seq:*
8: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[8].seq:*
9: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[9].seq:*
10: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[10].seq:*
11: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[11].seq:*
12: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[12].seq:*
13: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[13].seq:*
14: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[14].seq:*
15: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[15].seq:*
16: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[16].seq:*
17: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[17].seq:*
18: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[18].seq:*
19: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[19].seq:*
20: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[20].seq:*
21: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[21].seq:*
22: /cgn2_6/ptodata/1/ngeneseq/N_Geneseq_200[22].seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distributions.

SUMMARIES

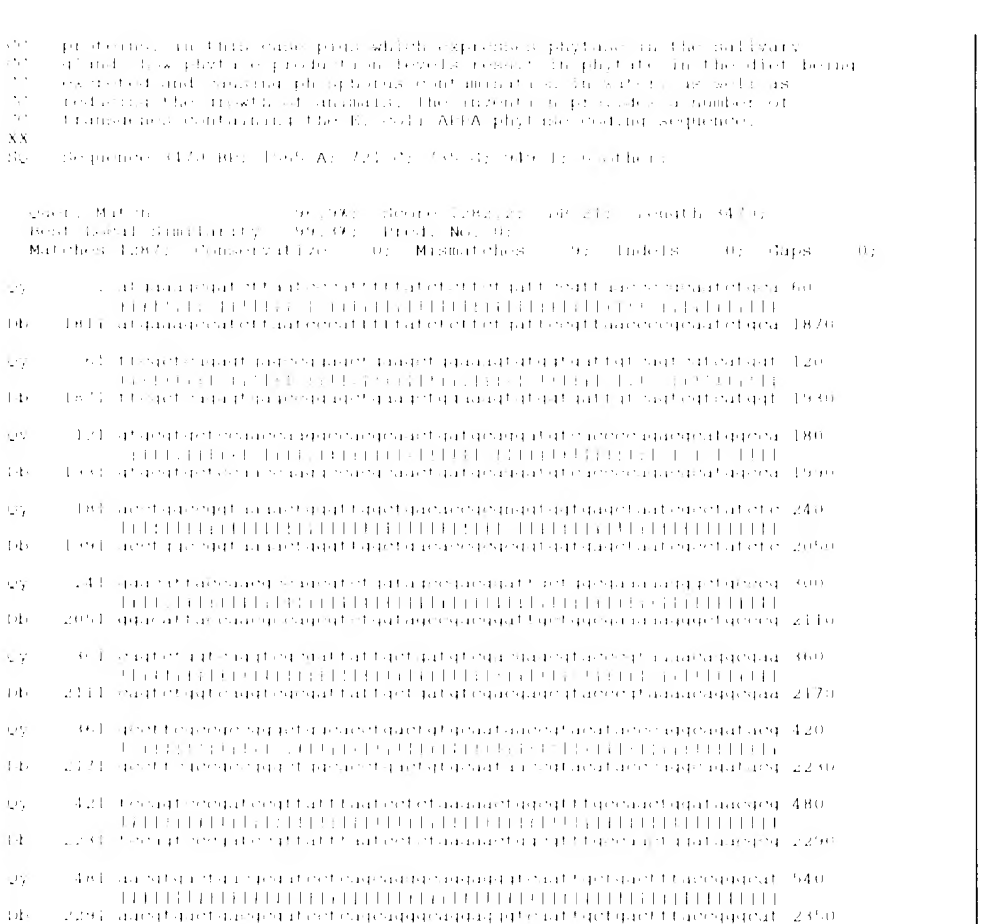
Result No. Score Match% Length DB ID Description

1	1423	100.0	1423	22	AAC88885	Escherichia coli p
c 2	1382.2	96.9	5427	21	AAC96268	bi%/AFPA plasmid c
c 3	1382.2	96.9	5427	21	AAC96270	bi%/AFPA plasmid c
4	1382.2	96.9	5427	21	AAC96267	bi%/AFPA plasmid c
5	1382.2	96.9	5427	21	AAC96269	bi%/AFPA plasmid c
6	1382.2	96.9	17734	21	AAC98000	Lumi2/AFPA plasmid
7	1382.2	96.9	20923	21	AAC96264	lumi2/AFPA plasmid
8	1375.2	96.6	5421	21	AAC96266	SV40/AFPA plasmid
9	1355.8	95.2	1160	21	AAA79236	F. coli acidic pho
10	1210	85.0	1272	20	AAX26540	DNA encoding an Es
c 11	86.2	6.1	427	21	AAA67371	Pinus radiata qlnc
c 12	81	5.6	998	19	AAV59100	Mouse nuclear fact
c 13	66.4	4.7	1901	21	AAA79864	Human ...
c 14	62	4.4	2511	18	AAB19553	Herpes ... DNA p
c 15	58	4.1	35166	20	AAV74802	KSHV I ... DNA (nuc
c 16	53	3.7	137507	19	AAV10941	KSHV I ... unique s
17	52.2	3.7	4427	11	AAQ06304	Sequence of plasmi
18	52.2	3.7	10732	21	AAA10594	Gene encoding a 50
19	51.8	3.6	11819	20	AAZ23905	Human ... DNA hexa
c 20	51.6	3.6	4209	21	AAC78080	Human tumor ... se
c 21	51.6	3.6	6466	18	AAB84206	mRNA for protein (
c 22	51.6	3.6	8470	21	AAZ95007	Cancer specific ge
c 23	51.6	3.6	8519	21	AAA37760	Human tumor antig
24	51.4	3.6	2890	19	AAX34100	HL ... ... DNA ...
c 25	51.2	3.6	1084	20	AAZ94429	ML ... ... an
c 26	51.2	3.6	1084	20	AAZ19217	ML ... ... re
27	51	3.6	1255	21	AAF12950	Aspergillus oryzae
28	51	3.6	4150	14	AAQ54128	Human ... DNA ...
29	51	3.6	4488	14	AAQ54126	Human ... DNA ...
30	51	3.6	4567	28	AAV43945	... ...
c 31	50.8	3.6	3273	21	AAA98710	... ...
c 32	50.8	3.6	22573	20	AAX37254	Human gene encodin
c 33	50.8	3.6	94618	21	AAF22285	BAC ... ...
c 34	50.6	3.6	94750	22	AAF28551	Genomic ... ... #
c 35	50.4	3.5	418	20	AAZ19419	ML ... ... an
c 36	50.4	3.5	418	28	AAZ19207	ML ... ... re
37	50.4	3.5	8252	15	AAQ69229	Host cell factor q
c 38	50.4	3.5	17882	20	AAX14166	... ...
39	50.4	3.5	46849	21	AAF22276	BAC ... ...
40	50.2	3.5	540	21	AAC74490	Human ... ... #
c 41	50.2	3.5	1316	18	AAB99543	Human ... ...
42	50.2	3.5	11968	21	AAF21105	Human ... ...
c 43	50.2	3.5	14068	21	AAF21272	Human ... ...
c 44	...	...	11968	21	AAA34983	Human ... ...
c 45	...	...	11968	21	AAA35150	Human ... ...

ALIGNMENTS

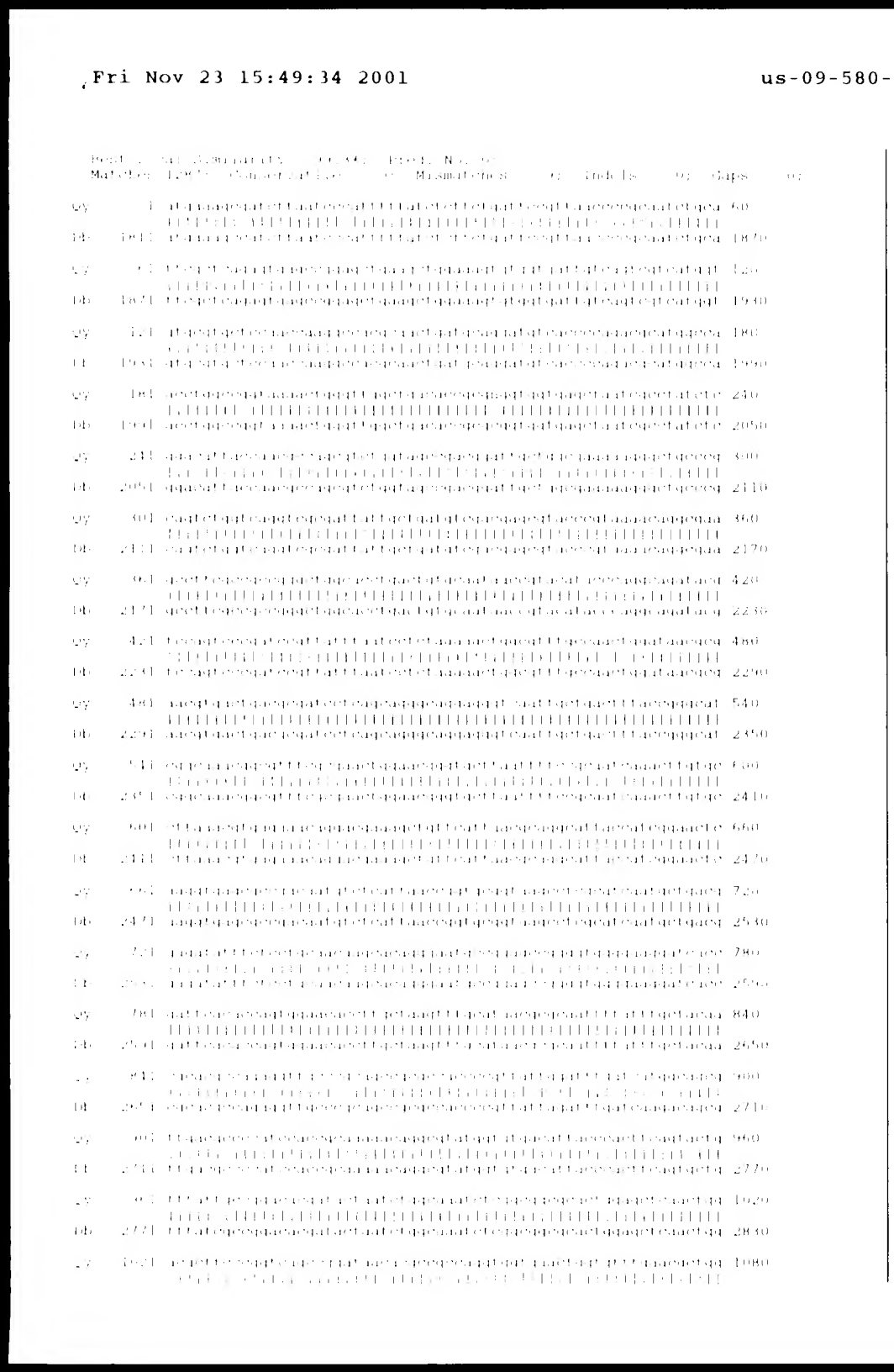
RESULT 1
AAC88885
ID AAC88885 standard; DNA; 1423 BP.
XX
AC AAC88885;
XX
DT ... (first entry)
XX
DE Escherichia coli ... phytase enzyme nucleotide sequence
XX
KW Escherichia coli; phytase enzyme; anabolic; phytate hydrolysis;
KW nutritional use.
XX
OS Escherichia coli.
XX
PN WO9621728-A1.
XX
PD 18-JUL-1996.
XX
PF 05-JAN-1996; 96WO-US000146.
XX
PR 05-JAN-1995; 95US-000369428.
XX
PA (DUPO) DUPONT CO.
XX
CC Start: 1, Stop: 1423, Strand: Plus
XX
DR P; P58822.
XX
FH Key Location/Qualifiers
FT source 1..1423
FT /organism="Escherichia coli"
FT /strain="AB1"
FT /plasmid="pAPPA"
FT /db_xref="taxon:562"
FT CDS 1..1423
FT /product="phytase enzyme"
FT /note="Improved the nutritional value of phytate containing feeds by
FT phytase enzymes which catalyze the liberation of inorganic phosphate
FT from the phytates"
XX

```
DE   Claim 2; Fig 1; 152pp; English.
XX
CC   The present sequence encodes a phytase enzyme from Escherichia coli B.
CC   The enzyme catalyses the liberation of inorganic phosphate from the
CC   phytate in phytate containing foodstuffs and can thus be used to improve
CC   the nutritional value of phytate rich ingredients.
XX
SQ   Sequence 1323 BP; 321 A; 351 C; 357 G; 289 T; 1 other.


  Query Match             99.9%;  Score 1322;  DB 2;  Length 1323;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1323;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgaaagcgatattaatcccattttttatctcttctgattccgttaaccccgcgaatctgca   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atgaaagcgatattaatcccattttttatctcttctgattccgttaaccccgcgaatctgca   60

Qy      61 ttcgctgatgatttgccccagagctgaagctgcagaagttgtcgatgttgattgtcagttcgctgatggt  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ttcgctgatgatttgccccagagctgaagctgcagaagttgtcgatgttgattgtcagttcgctgatggt  120

Qy     121 atcgatgctgtcgtcgtcagccacgcgtatgtcgcgcagtatgtcgaacgcgatcagcgcgcg  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 atcgatgctgtcgtcgtcagccacgcgtatgtcgcgcagtatgtcgaacgcgatcagcgcgcg  180

Qy     181 acctatcgcgattaaaactgattggtcgatcaccgcgaagttcgtcgcactatcgccgatatctc  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 acctatcgcgattaaaactgattggtcgatcaccgcgaagttcgtcgcactatcgccgatatctc  240

Qy     241 tgatattaccgaacgccatcgtctgatcgtcctcgcatcgtattgctgccgtaaaacaggctgccg  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 tgatattaccgaacgccatcgtctgatcgtcctcgcatcgtattgctgccgtaaaacaggctgccg  300

Qy     301 cattctgatgcgaatatcgccattattgctgatgtgcgcagcgatctgccgctgaaagcgctgcag  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 cattctgatgcgaatatcgccattattgctgatgtgcgcagcgatctgccgctgaaagcgctgcag  360

Qy     361 gcctttcgcgccgggctgaccttgctgcatgccgcgctgaccgtgccgcagcgtatgcg  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gcctttcgcgccgggctgaccttgctgcatgccgcgctgaccgtgccgcagcgtatgcg  420

Qy     421 tccgatcgcgatatcattattaatctcctcgacgactggcgtgattgccgcactgcgtatgacgcgt  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tccgatcgcgatatcattattaatctcctcgacgactggcgtgattgccgcactgcgtatgacgcgt  480

Qy     481 aacgtgacttaccgcgatctcaccgacgagcgcgtgaactgcattgctgacttaccgggcgcat  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 aacgtgacttaccgcgatctcaccgacgagcgcgtgaactgcattgctgacttaccgggcgcat  540

Qy     541 cgtcaggcgcgtcagttcgccgacctgcgcgaccgggtgcttgatttccgccgcatcgaacttgtgc  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 cgtcaggcgcgtcagttcgccgacctgcgcgaccgggtgcttgatttccgccgcatcgaacttgtgc  600

Qy     601 cttgaacgtttgtatcgcgatgatcgtgatgcgttcgatgcgcgtgatttcgatgcgaacgaa  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 cttgaacgtttgtatcgcgatgatcgtgatgcgttcgatgcgcgtgatttcgatgcgaacgaa  660

Qy     661 aacgtgaagcgcgtcgcgaatatgtcgaattacaggaagatggcgaattcgtcgtgctgtcgctgcc  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 aacgtgaagcgcgtcgcgaatatgtcgaattacaggaagatggcgaattcgtcgtgctgtcgctgcc  720

Qy     721 tgatattttcctcgtcacgaagcgcagcgtgatcgcgagcgtcgtgtgcgcgaagatgcgctgcc  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 tgatattttcctcgtcacgaagcgcagcgtgatcgcgagcgtcgtgtgcgcgaagatgcgctgcc  780

Qy     781 gattgcaccattgaacacctttgctaagttgcataacgcgcgcattttatttgctgcaa  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 gattgcaccattgaacacctttgctaagttgcataacgcgcgcattttatttgctgcaa  840

Qy     841 cgcagccgcagaagttgccgcagccgcgccgccaccccgttattgaattgaatcatgcagccg  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 cgcagccgcagaagttgccgcagccgcgccgccaccccgttattgaattgaatcatgcagccg  900

Qy     901 ttaacgcgcgcgcatccgcgcgcaaaacagcgcgtatgatgatgcgacattaccgcgcacttgcatgc  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ttaacgcgcgcgcatccgcgcgcaaaacagcgcgtatgatgatgcgacattaccgcgcacttgcatgc  960

Qy     961 tttattgccgaacacgatgctaatctgacgaaatctgccgccgcgctgcaactgaagctcaactaa  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 tttattgccgaacacgatgctaatctgacgaaatctgccgccgcgctgcaactgaagctcaactaa  1020

Qy    1021 acgctttccgcgtcagccgaataacacgccgccaatacgtgaactcgttattttgcgtactcgt  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 acgctttccgcgtcagccgaataacacgccgccaatacgtgaactcgttattttgcgtactcgt  1080

Qy    1081 cgtcgcctaaagccgataacagccagtcgattcagggtttcgctgcgtctttccagacgtttacgag  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 cgtcgcctaaagccgataacagccagtcgattcagggtttcgctgcgtctttccagacgtttacgag  1140

Qy    1141 cagatacgtcgtgatacaaacgccgctgtcattaaatacgccgccgccgataatgtgaacctgacc  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 cagatacgtcgtgatacaaacgccgctgtcattaaatacgccgccgccgataatgtgaacctgacc  1200

Qy    1201 ctgccagcgatgtcgatgacgccgacgtcgcagcgccgcatatgttcgttcgttacgccagtttta  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ctgccagcgatgtcgatgacgccgacgtcgcagcgccgcatatgttcgttcgttacgccagtttta  1260

Qy    1261 atcgtgaatgaagcacgcgatacccgcgctgcgaatttgaagatctcatcaccgatcaccatcgtcc  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 atcgtgaatgaagcacgcgatacccgcgctgcgaatttgaagatctcatcaccgatcaccatcgtcc  1320

Qy    1321 taa  1323
           |||
Db    1321 taa  1323


RESULT   2
AAV68298
ID   AAV68298 standard; DNA; 5470 BP.
XX
AC   AAV68298;
XX
DT   20-FEB-2001  (first entry)
XX
DE   pIG/AfiA plasmid coding sequence.
XX
KW   transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW   environmental pollution; pig; ds.
XX
OS   Battus sp.  chimeric.
OS   Escherichia coli  chimeric.
XX
PN   WO200064247-A1.
XX
PD   02-NOV-2000.
XX
PF   20-APR-2000;  2000WO-CA00430.
XX
PR   23-APR-1999;  1999US-0130508.
XX
PA   (GYOS ) UNIV GUELPH.
XX
PI   Forsberg CW;  Golovan S;  Phillips JP;
XX
DR   WPI; 2000-672219/67.
DR   P-PSDB; AAB36261.
XX
PT   Transgenic non-human animal for gastrointestinal tract specific
PT   expression of a protein, preferably phytase, comprises a nucleic acid
PT   sequence including a heterologous transgene construct encoding the
PT   protein -
XX
PS   Claim 14; Fig 21; 152pp; English.
XX
CC   The present invention provides transgenic animals which p ss s  s ired
```

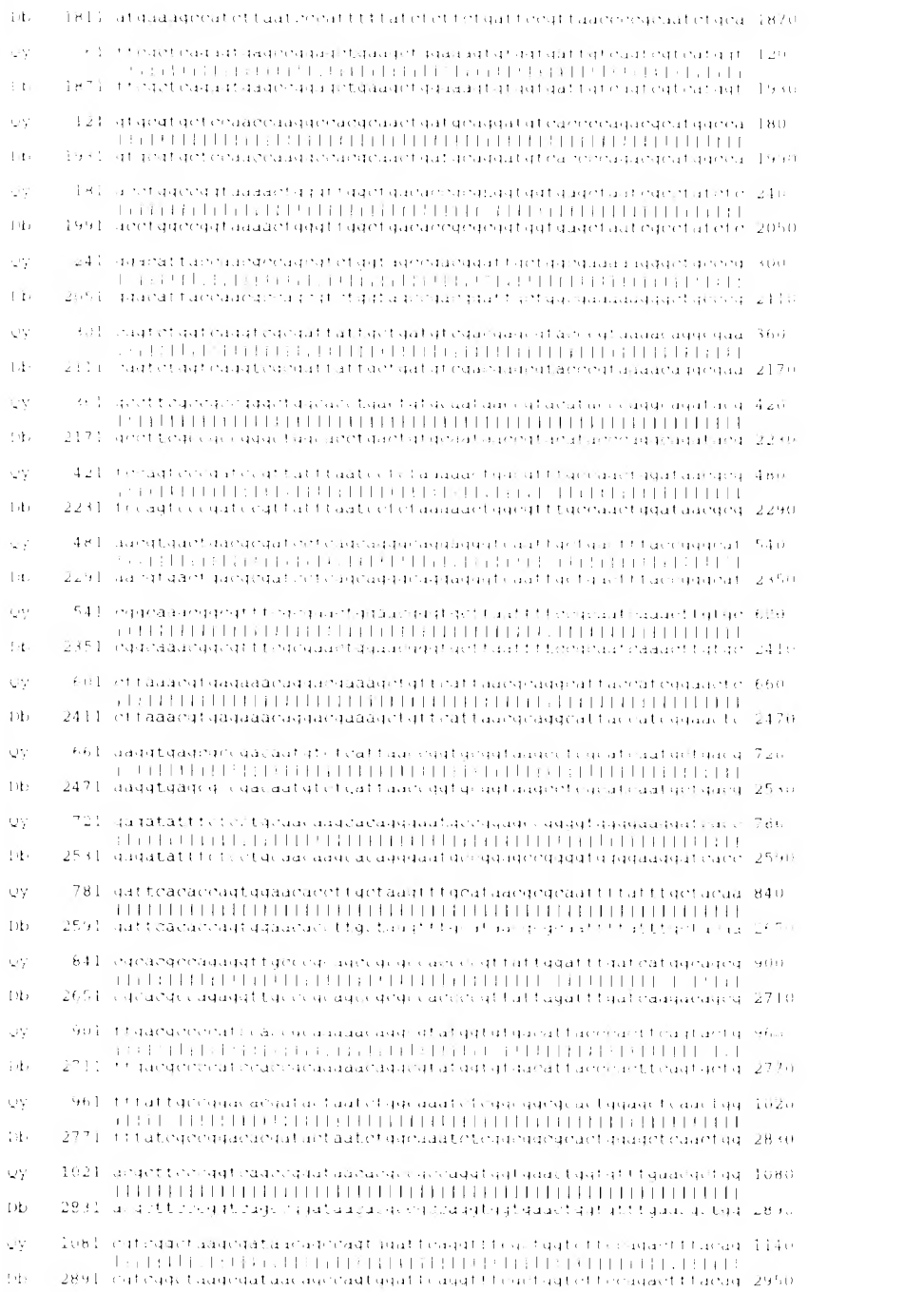


Result 4
AAO68296
ID AAO68296 standard; DNA; 4060 BP.
XX
AC AAO68296;
XX
DT 20-FEB-2001 (first entry)
XX
DE RF5/APPA plasmid coding sequence.
XX
KW transgenic animal; salivary protein; phytase; phosphorus; waste; water; environmental pollution; pig; ds.
XX
OS Rattus sp. chimeric.
OS Escherichia coli chimeric.
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA00463.
XX
PR 23-APR-1999; 99US-0130508.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI; 2000-667245/67.
DR P-PSDB; AAB36259.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein
XX
PS Claim 11; Fig 1a; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence

```
XX
SQ   Sequence 4060 BP; 1257 A; 814 C; 843 G; 1146 T; 0 other;

  Query Match             96.9%;  Score 1282.2;  DB 21;  Length 4060;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 1287;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 atgaaagcgatcttaatcccatttttatctcttctgattccgttaacccgcaatctgca 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1811 atgaaagcgatcttaatcccatttttatctcttctgattccgttaacccgcaatctgca 1870

Qy         61 ttcgctcagagtgagccggagctgaagctggaaagtgtggtgattgtcagtcgtcatggt 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1871 ttcgctcagagtgagccggagctgaagctggaaagtgtggtgattgtcagtcgtcatggt 1930

Qy        121 gtgcgtgctccaaccaaggccacgcaactgatgcaggatgtcaccccagacgcatggcca 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1931 gtgcgtgctccaaccaaggccacgcaactgatgcaggatgtcaccccagacgcatggcca 1990

Qy        181 acctggccggtaaaactgggttggctgacaccgcgcggtggtgagctaatcgcctatctc 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1991 acctggccggtaaaactgggttggctgacaccgcgcggtggtgagctaatcgcctatctc 2050

Qy        241 ggacattaccaacgccagcgtctggtagccgacggattgctggcgaaaaagggctgcccg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2051 ggacattaccaacgccagcgtctggtagccgacggattgctggcgaaaaagggctgcccg 2110

Qy        301 cagtctggtcaggtcgcgattattgctgatgtcgacgagcgtacccgtaaaacaggcgaa 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2111 cagtctggtcaggtcgcgattattgctgatgtcgacgagcgtacccgtaaaacaggcgaa 2170

Qy        361 gccttcgccgccgggctggcacctgactgtgcaataaccgtacatacccaggcagatacg 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2171 gccttcgccgccgggctggcacctgactgtgcaataaccgtacatacccaggcagatacg 2230

Qy        421 tccagtcccgatccgttatttaatcctctaaaaactggcgtttgccaactggataacgcg 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2231 tccagtcccgatccgttatttaatcctctaaaaactggcgtttgccaactggataacgcg 2290

Qy        481 aacgtgactgacgcgatcctcagcagggcaggagggtcaattgctgactttaccgggcat 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2291 aacgtgactgacgcgatcctcagcagggcaggagggtcaattgctgactttaccgggcat 2350

Qy        541 cggcaaacggcgtttcgcgaactggaacgggtgcttaattttccgcaatcaaacttgtgc 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2351 cggcaaacggcgtttcgcgaactggaacgggtgcttaattttccgcaatcaaacttgtgc 2410

Qy        601 cttaaacgtgagaaacaggacgaaagctgttcattaacgcaggcattaccatcggaactc 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2411 cttaaacgtgagaaacaggacgaaagctgttcattaacgcaggcattaccatcggaactc 2470

Qy        661 aaggtgagcgccgacaatgtctcattaaccggtgcggtaagcctcgcatcaatgctgacg 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2471 aaggtgagcgccgacaatgtctcattaaccggtgcggtaagcctcgcatcaatgctgacg 2530

Qy        721 gagatatttctcctgcaacaagcacagggaatgccggagccggggtggggaaggatcacc 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2531 gagatatttctcctgcaacaagcacagggaatgccggagccggggtggggaaggatcacc 2590

Qy        781 gattcacaccagtggaacaccttgctaagtttgcataacgcgcaatttttatttgctacaa 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2591 gattcacaccagtggaacaccttgctaagtttgcataacgcgcaatttttatttgctacaa 2650

Qy        841 cgcacgccagaggttgcccgcagccgcgccaccccgttattagatttgatcatggcagcg 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2651 cgcacgccagaggttgcccgcagccgcgccaccccgttattagatttgatcaaggcagcg 2710

Qy        901 ttgacgccccatccacctgcaaaacaagcgtatggtgtgacattacccacttcagtgctg 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2711 ttgacgccccatccacctgcaaaacaagcgtatggtgtgacattacccacttcagtgctg 2770

Qy        961 tttattgccggacacgatactaatctggcaaatctcggcggcgcactggagctcaactgg 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2771 tttatcgccggacacgatactaatctggcaaatctcggcggcgcactggagctcaactgg 2830

Qy       1021 acgcttcccggtcagccggataacacgccgccaggtggtgaactggtgttttgaacgctgg 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2831 acgcttcccggtcagccggataacacgccgccaggtggtgaactggtgttttgaacgctgg 2890

Qy       1081 cgtcggctaagcgataacagccagtggattcaggtttcgctggtcttccagactttacag 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2891 cgtcggctaagcgataacagccagtggattcaggtttcgctggtcttccagactttacag 2950

Qy       1141 cagatgcgtgataaaacgccgctgtcattaaatacgccgcccggagaggtgaaactgacc 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2951 cagatgcgtgataaaacgccgctgtcattaaatacgccgcccggagaggtgaaactgacc 3010

Qy       1201 ctggcaggatgtgaagagcgaaatgcgcagggcatgtgttcgttggcaggtttacgcaa 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3011 ctggcaggatgtgaagagcgaaatgcgcagggcatgtgttcgttggcaggtttacgcaa 3070

Qy       1261 atcgtgaatgaagcacgcataccggcgtgcagtttg 1296
              ||||||||||||||||||||||||||||||||||||
Db       3071 atcgtgaatgaagcacgcataccggcttgcagtttg 3106


RESULT 4
AAC68297
ID   AAC68297 standard; DNA; 6116 BP.
XX
AC   AAC68297;
XX
DT   26-FEB-2001  (first entry)
XX
DE   KP5/APPA plasmid coding sequence.
XX
KW   Transgenic animal; salivary protein; phytase; phosphorus; associated with
KW   environmental pollution; pig; ds.
XX
OS   Rattus sp.  chimeric.
OS   Escherichia coli  chimeric.
XX
PN   WO200064247-A1.
XX
PD   02-NOV-2000.
XX
PF   20-APR-2000; 2000WO-CA00430.
XX
PR   23-APR-1999;   99US-0130508.
XX
PA   (GUEL-) UNIV GUELPH.
XX
PI   Forsberg CW,  Golovan S,  Phillips JP;
XX
DR   WPI; 2000-672457/65.
DR   P-PSDB; AAB67260.
XX
PT   Transgenic non-human animal for gastrointestinal tract specific
PT   expression of a protein, preferably phytase, comprises a nucleic acid
PT   sequence including a heterologous transgene construct encoding the
PT   protein
XX
PS   Claim 56; Fig 20; 152pp; English.
XX
CC   The present invention provides transgenic animals which produce desired
CC   proteins, in this case pigs which expresses phytase in the salivary
CC   gland. Low phytase production levels result in phytate in the diet being
CC   excreted and causing phosphorus contamination in water, as well as
CC   reducing the growth of animals. The invention provides a number of
CC   transgenes containing the E. coli APPA phytase coding sequence.
XX
SQ   Sequence 6116 BP; 1724 A; 1386 C; 1407 G; 1599 T; 0 other;

  Query Match             96.9%;  Score 1282.2;  DB 21;  Length 6116;
```



RESULT 5
AAC68295
ID AAC68295 standard; DNA; 6708 BP.
XX
AC AAC68295;
XX
DT 20-FEB-2001 (first entry)
XX
DE 81%/APPA phased coding sequence.
XX
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig; ds.
XX
OS Ratus sp. - chimeric.
OS Escherichia coli - chimeric.
XX
RN W:200064247 A1.
XX
RD 02-NOV-2000.
XX
RF 20-APR-2000; 2000WO-CA00430.
XX
PR 23-APR-1999; 99US-0130508.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI; 2000-687245/67.
DR P-PSDB; AAB36258.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a vector and
PT sequence including a heterologous transgene construct encoding the
PT protein
XX
PS Claim 56; Fig 18; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence.
XX
SQ Sequence 6708 BP; 1916 A; 1479 C; 1515 G; 1798 T; 0 other;



Qy 1141 caaatgcgtgatgaaaagccgctgtcattaaatacgccgccgaaagaatctaccgc 1200
Db 2951 caaatgcgtgatgaaaagccgctgtcattaaatacgccgccgaaagaatctaccgc 3010

Qy 1201 ctgcgcgatgtgtgaaagaaggaatgccgcgaggcatgtgttcgttgacgagttttcg 1260
Db 3011 ctgacgaggatgtgaaagaaggaatgccgcgaggcatgtgttcgttgacgagttttcg 3070

Qy 1261 atcgtcaatgaagcacgcatgcgcgctgcgagttta 1296
Db 3071 atcgtgaatgaagcacgcatgcgcgctgcgagttta 3106

RESULT 9
AAC08800
ID AAC08800 standard; DNA; 1732 BP.
XX
AC AAC08800;
XX
DT 20-FEB-2001 (first entry)
XX
DE rumen27APPA plasmid coding sequence.
XX
KW transgenic animal; salivary protein; phytase; phosphatase; animal nutrition;
KW environmental pollution; pig; ds.
XX
OS Mus musculus - chimeric.
OS Escherichia coli - chimeric.
XX
RN [1]
RP 1-1732
XX
RA Forsberg CW, Golovan S, Phillips JP;
XX
RT "Transgenic non human animal for gastrointestinal tract specific
RT expression of a protein, preferably phytase, comprises a nucleic acid
RT sequence including a heterologous transgene construct encoding the
RT protein";
XX
RL Patent number WO200064247-A1, 02-NOV-2000.
XX
RF 28-APR-2000; 2000WO-CA000490.
XX
PR 29-APR-1999; 99US-0131058P.
XX
PA (UYGU-) UNIV SASKATCHEWAN.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI; 2000-683245/67.
DR P-PSDB; AAB06263.
XX
PT Transgenic non human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein -
XX
PS Claim 14; Fig 23; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence.
XX
SQ Sequence 1732 BP; 471G A; 412C C; 416G G; 431T T; 1 other;

Query Match 96.9%; Score 1282.2; DB 21; Length 1732;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgaaagcgatcttaatccattttatctcttctgattcgttaccccgcaatctcaa 60
Db 1266 atgaaagcgatcttaatccattttatctcttctgattcgttaccccgcaatctcaa 1325

Qy 61 ttcgctgaaagtgagcctgagctgaagctgcagtcggtgagcgataccggtgtgcag 120
Db 1326 ttcgctgaaagtgagcctgagctgaagctgcagtcggtgagcgataccggtgtgcag 1385

Qy 121 gtgcagctgctggaagacagcgaacaggcggacaagctgatccagcagatctatgtcgtc 180
Db 1386 gtgcagctgctggaagacagcgaacaggcggacaagctgatccagcagatctatgtcgtc 1445

Qy 181 agccatggcgtgcgtagcccgctgaccgatgaacagacgcagtggatcgaggaccgtctg 240
Db 1446 agccatggcgtgcgtagcccgctgaccgatgaacagacgcagtggatcgaggaccgtctg 1505

Qy 241 gcgaaagatgccgaaaaccgtcatcctgatccgcgcgcaaggtggccgcgaaaccgtgcg 300
Db 1506 gcgaaagatgccgaaaaccgtcatcctgatccgcgcgcaaggtggccgcgaaaccgtgcg 1565

Qy 301 caactgatgcaagtgtttgagcgcattattgctgatgtggatagcgcgctgaccgaagag 360
Db 1566 caactgatgcaagtgtttgagcgcattattgctgatgtggatagcgcgctgaccgaagag 1625

Qy 361 gcggtgccgcagctgaaagcgctgaaaaacagcgcgctggatgtgggcattgcgtatgag 420
Db 1626 gcggtgccgcagctgaaagcgctgaaaaacagcgcgctggatgtgggcattgcgtatgag 1685

Qy 421 accagcgtgaaagatgatgcggtgagcttcaacaacctgagcagcaccctgccgattagc 480
Db 1686 accagcgtgaaagatgatgcggtgagcttcaacaacctgagcagcaccctgccgattagc 1745

Qy 481 gaagtgaatgtgaccgaagatgaagatgtgctgcatggctatccgtttctgaaagatccg 540
Db 1746 gaagtgaatgtgaccgaagatgaagatgtgctgcatggctatccgtttctgaaagatccg 1805

Qy 541 cgcgcgattcgccagtttctgcagcgctgcgcgctgtttcgcgcgattgaaaacctgcgc 600
Db 1806 cgcgcgattcgccagtttctgcagcgctgcgcgctgtttcgcgcgattgaaaacctgcgc 1865

Qy 601 cttaaacgtgcgaaagcgtatcgcgaagtgattccgattaaccgcgcgctgattaccatc 660
Db 1866 cttaaacgtgcgaaagcgtatcgcgaagtgattccgattaaccgcgcgctgattaccatc 1925

Qy 661 gaaattgaagcgcgcgaagatgtgctgattaacgaatggcgcgatgaaaccctgcagatc 720
Db 1926 gaaattgaagcgcgcgaagatgtgctgattaacgaatggcgcgatgaaaccctgcagatc 1985

Qy 721 gatatattcctgctgaccaaccaccaggatgaccgcgaaattaccgaagtgaaagcgatc 780
Db 1986 gatatattcctgctgaccaaccaccaggatgaccgcgaaattaccgaagtgaaagcgatc 2045

Qy 781 gattcacaccaatgaagcacgtttgctgaatttgcgtatgaacgcgaattttatttgctc 840
Db 2046 gattcacaccaatgaagcacgtttgctgaatttgcgtatgaacgcgaattttatttgctc 2105

Qy 841 cgcaacgcgaaagaatgccgcgcagcagcgcgcgcaccgttatttgaatttgatcatcgc 900
Db 2106 cgcaacgcgaaagaatgccgcgcagcagcgcgcgcaccgttatttgaatttgatcatcgc 2165

Qy 901 ttaaacgtgcgtatcaccgcgaaagaagtggatgtgattgaaattaccgaagtgctgcag 960
Db 2166 ttaaacgtgcgtatcaccgcgaaagaagtggatgtgattgaaattaccgaagtgctgcag 2225

Qy 961 tttattgcgcgcaacgtgtatctgaatccgaatctgcaggaagcgctgaacggcgtgacc 1020
Db 2226 tttattgcgcgcaacgtgtatctgaatccgaatctgcaggaagcgctgaacggcgtgacc 2285

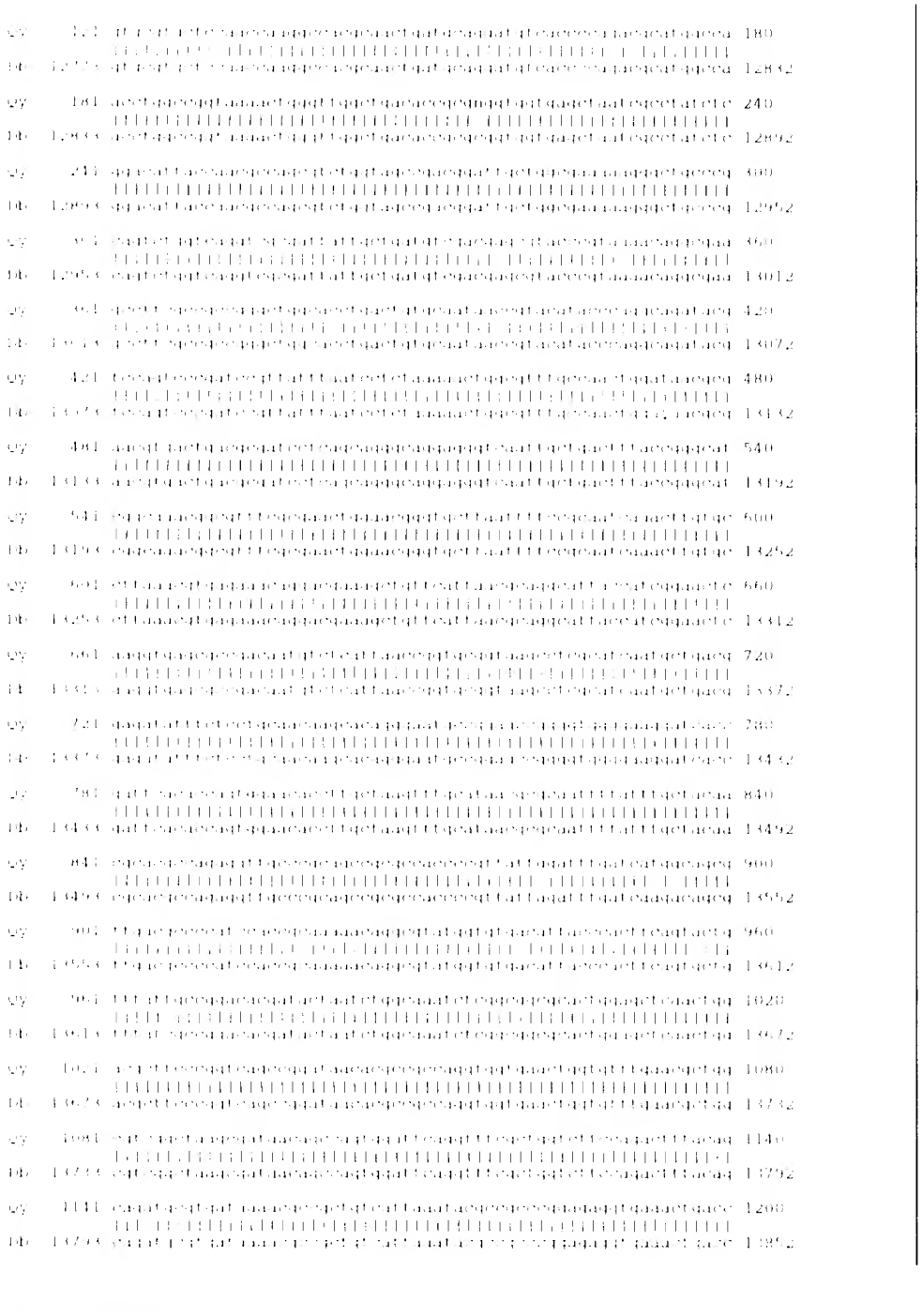
Qy 1021 acgctgtcgcagctgtttgcgagcaacatgcagcagattattggcgcggtgaaa 1080
Db 2286 acgctgtcgcagctgtttgcgagcaacatgcagcagattattggcgcggtgaaa 2345

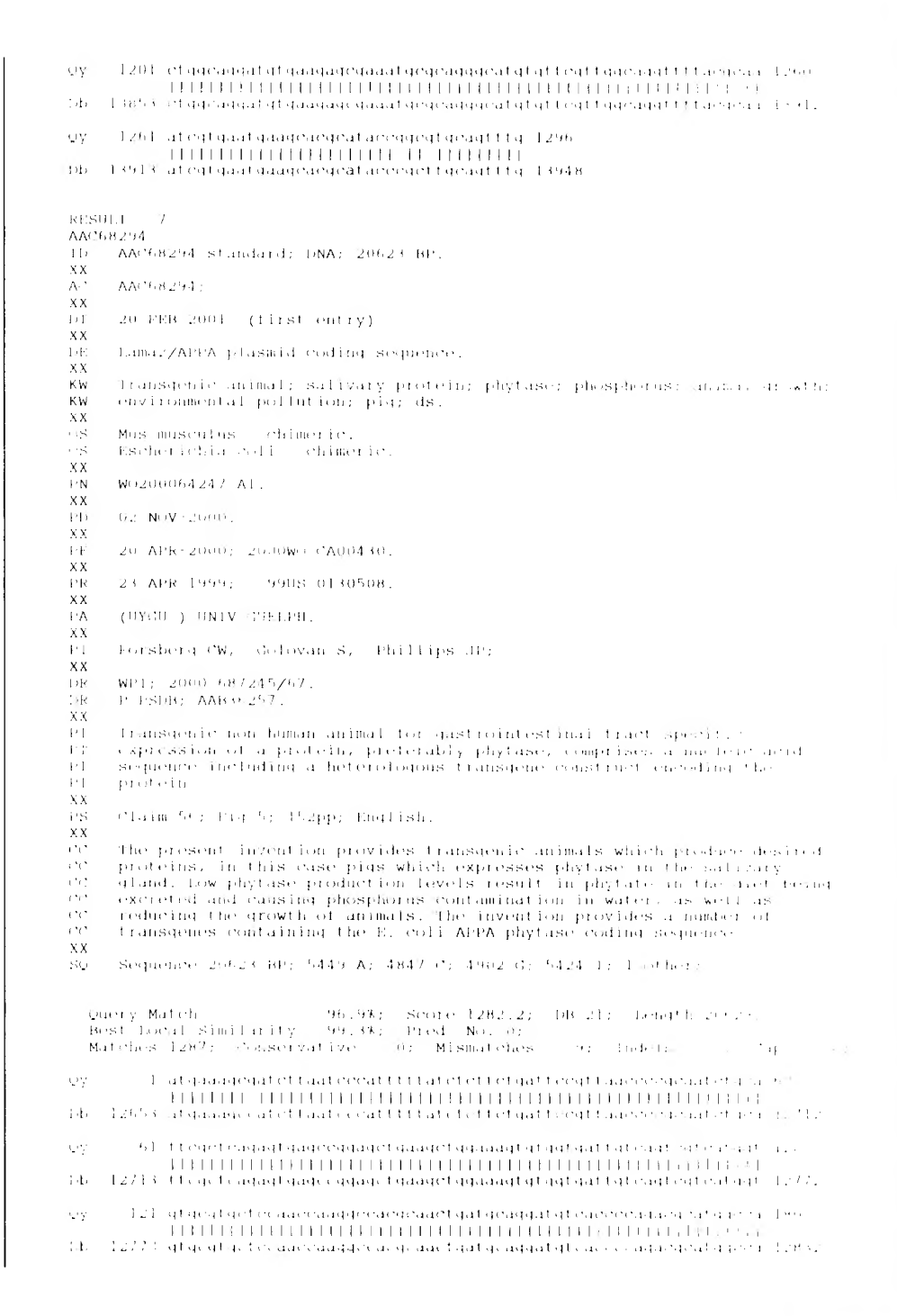
Qy 1081 ctgcgcctgaccgatcaccagattgtgattccgagcagcctgccgagcattaccgaa 1140
Db 2346 ctgcgcctgaccgatcaccagattgtgattccgagcagcctgccgagcattaccgaa 2405

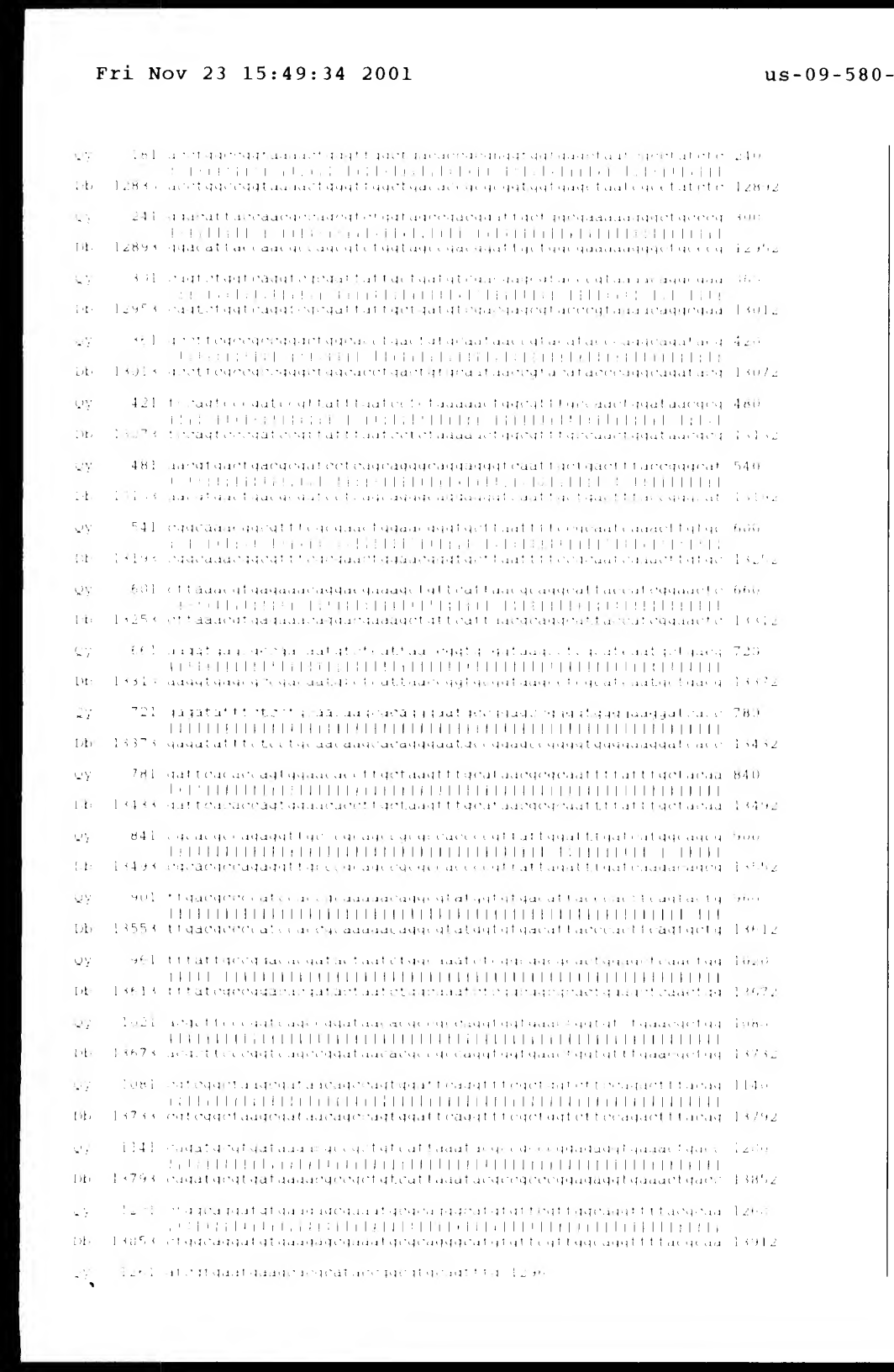
Qy 1141 caagtgcgtgatgaaaagccgctgtcattaaatacgccgccgaaagaatctaccgc 1200
Db 2406 caagtgcgtgatgaaaagccgctgtcattaaatacgccgccgaaagaatctaccgc 2465

Qy 1201 ctgcgcgatgtgtgaaagaaggaatgccgcgaggcatgtgttcgttgacgagttttcg 1260
Db 2466 ctgacgaggatgtgaaagaaggaatgccgcgaggcatgtgttcgttgacgagttttcg 2525

Qy 1261 atcgtcaatgaagcacgcatgcgcgctgcgagttta 1296
Db 2526 atcgtgaatgaagcacgcatgcgcgctgcgagttta 2561



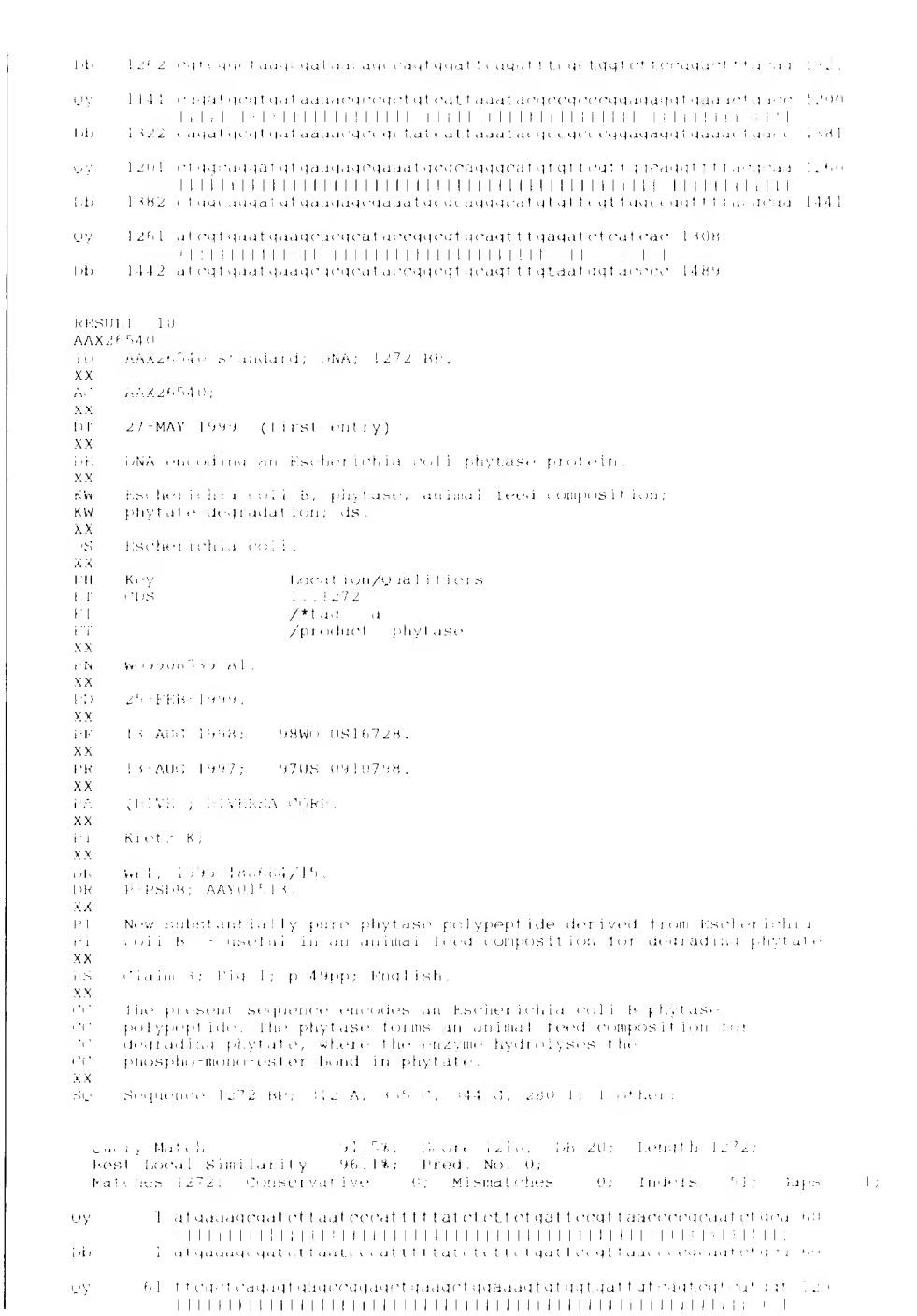


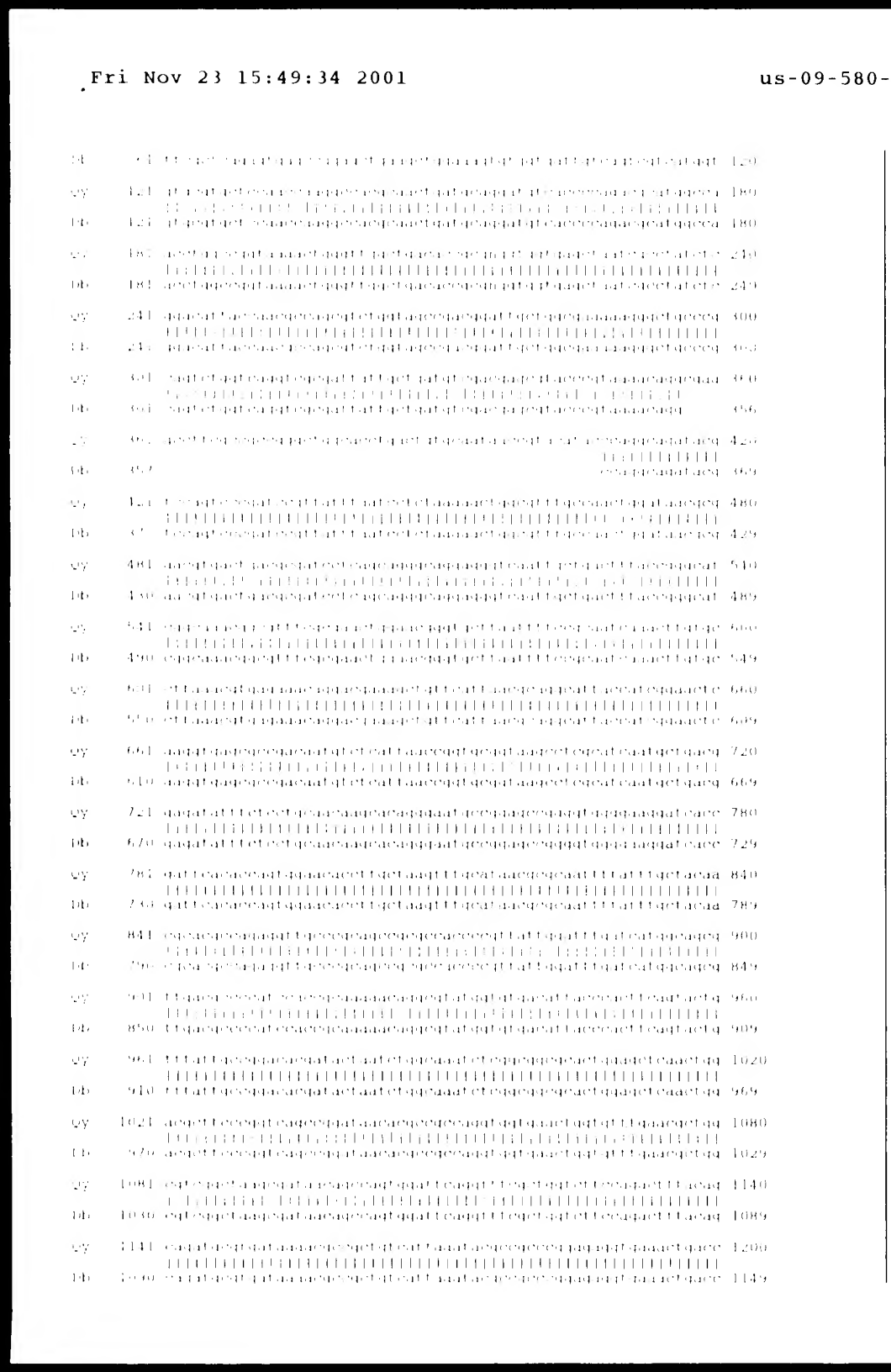


RESULT 8
AAC68299
ID AAC68299 standard; DNA; 5421 BP.
XX
AC AAC68299;
XX
DT 20-FEB-2001 (first entry)
XX
DE AVIc/AEA plasmid coding sequence.
XX
KW transgenic animal; salivary protein; phytase; phosphorus; assay; growth;
KW environmental pollution; pig; ds.
XX
OS Rhesus macaque polyomavirus - chimeric.
OS Escherichia coli - chimeric.
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA00730.
XX
PR 28-APR-1999; 99US-0131098.
XX
PA (UYGH-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI; 2000-706724/67.
DR P-PSDB; AAB39262.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a gene, e.g. for, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein
XX
PS Claim 56; Fig 22; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expressos phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in waste, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence.
XX
SQ Sequence 5421 BP; 1413 A; 1321 C; 1331 G; 1355 T; 1 other;

Query Match 96.8%; Score 1270.2; DB 21; Length 5421;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1297; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

AAA28216
ID AAA28216 standard; DNA; 1489 BP.
XX
AC AAA28216;
XX
DT 12-FEB-2001 (first entry)
XX
DE E. coli acidic phosphatase appA2 encoding nucleotide sequence.
XX
KW Phosphatase; phytase activity; protease; acidic phosphatase; appA;
KW periplasmic phosphoanhydride phosphohydrolase; animal feed.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT protein_bind complement (1..22)
FT /*tag= a
FT /bound_moiety= "Primer P11"
FT CDS 16..108
FT /*tag= b
FT /product= "AppA2 related peptide"
FT CDS 184..1482
FT /*tag= c
FT /product= "AppA2"
FT /note= "Acidic phosphatase"
FT primer_bind 241..261
FT /*tag= d
FT /bound_moiety= "Primer K2"
FT primer_bind 1471..1489
FT /*tag= e
FT /bound_moiety= "Primer K2"
XX
PN WO200058481-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08590.
XX
PR 31-MAR-1999; 99US-0127032.
XX
PA (CORR-) CORNELL RES FOUND INC.
XX
PI Lei X;
XX
DR WPI; 2000-611479/59.
DR P-PSDB; AAY94753, AAY94756.
XX
PT Novel phosphatase having improved phytase activity useful in animal
PT feed for improving the accessibility of phosphate to an animal is
PT produced by treating phosphatase with protease
XX
PS Disclosure; Fig 6; 48pp; English.
XX
CC This invention relates to a phosphatase fragment with improved phytase
CC activity. The phosphatase fragment is created by treating the phosphatase
CC with a protease. The invention includes an Escherichia coli-derived
CC sequence which encodes an acidic phosphatase appA2 which has improved
CC phosphatase activity. The appA gene of Escherichia coli was originally
CC defined as E. coli periplasmic phosphoanhydride phosphohydrolase gene,
CC although it was not previously known to have phytase activity. The
CC phosphatase fragment can be used in animal feed, to improve the
CC accessibility of phosphate to the animal. The phosphatase gene which
CC encodes the protein can be used in the production of phytase. The present
CC sequence represents the appA2 coding sequence.
XX
SQ Sequence 1489 BP; 362 A; 384 C; 403 G; 340 T; 0 other;





RESULT 11
AAA67371/c
ID AAA67371 standard; DNA; 427 BP.
XX
AC AAA67371;
XX
DT 31-OCT-2000 (first entry)
XX
DE Pinus radiata glucose 6 phosphate dehydrogenase DNA Fig 1c SEQ ID NO 3.
XX
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; ds.
XX
OS Pinus radiata.
XX
FH W 2000/209... A..
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339428/29.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant
XX
PS Claim 1; Page 170; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 845 nucleotide sequences (see AA
CC AAA67673 to AAA67967, their (reverse) complements, sequences at least
CC an Expectation (E) value of 0.01 or less compared to the 845 sequences,
CC sequences at least 50% identical to them, 200, 150, 40 or 20 nucl. (all
CC 845 sequences or sequences that are degenerately equivalent. A number
CC to the 845 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAM15208
CC to AAM15840 are proteins encoded by some of the polynucleotide sequences
CC given in the present invention.
XX
SQ Sequence 427 BP; 85 A; 118 C; 114 G; 109 T; 2 other;

Query Match 2.7%; Score 36.2; DB 21; Length 1;
Best Local Similarity 47.3%; Pred. No. 0.12;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

RESULT 12
AAV59[illegible
ID AAV59[illegible standard; cDNA; 998 BP.
XX
AC AAV59[illegible;
XX
DT 2[illegible JAN 1999 (first entry)
XX
DE Mouse amphiregulin coding sequence.
XX
KW Amphiregulin; AR protein; epithelial cell growth; tumour cell growth; psoriasis; tumour marker; cancer diagnosis; therapy; mouse; ss.
XX
OS Mus sp.
XX
PN [illegible] OF A
XX
PD 08 NOV 1998.
XX
PF 18-MAY-1992; 92JP-0885089.
XX
PR 18-MAY-1992; 92JP-0885089.
PR 25-JAN-1988; 88US-0148427.
PR 15-APR-1988; 88US-0181884.
PR 17-JAN-1989; 89US-0297816.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Bradley JG, McDonald VL, Plowman GD, Shoyab M,
XX
DR WPI; 1998-[illegible]-51.
DR P-PSDB; AAW74162.
XX
PT Amphiregulin proteins useful for stimulating e.g. epithelial cell growth.
XX
PS Example; Fig 1c; 79pp; English.
XX
CC This sequence encodes the mouse amphiregulin (AR) of the invention.
CC The invention also relates to rat, bovine and human AR proteins. The AR
CC proteins can be used to stimulate epithelial cell growth. They may be of
CC therapeutic use in controlling tumour cell growth and in disorders such
CC as psoriasis, and can be used as a tumour marker for cancer diagnosis.
XX
SQ Sequence 998 BP; 320 A; 188 C; 241 G; 249 T; 0 other;

Query Match 2.6%; Score 34; DB 19; Length 998;
Best Local Similarity 5.1%; Pred. No. 1;
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

RESULT 13
AAC79864/c
ID AAC79864 standard; cDNA; 1901 BP.
XX
AC AAC79864;
XX
DT 09-FEB-2001 (first entry)
XX
DE Human secreted protein encoding cDNA for gene 16.
XX
KW human; secreted protein; cytostatic; antiarthritic; antiasthmatic;
KW immunosuppressive; antiatherosclerotic; antiinflammatory; neuroprotective;
KW antidiabetic; tranquiliser; vulnerary; antibacterial;
KW antipsoriatic; antiarrhythmic; antirheumatic; cardiant; anti HIV;
KW autoimmune disorder, allergic condition, cardiovascular disorder;
KW cancer; neurological disease; tissue repair; ss.
XX
OS Homo sapiens.
XX
PN WO200055176-A2.
XX
PD 21-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06057.
XX
PR 12-MAR-1999; 99US-0124142.
PR 11-JUN-1999; 99US-0138597.
PR 03-DEC-1999; 99US-0168666.
XX
PA (HUMA) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-[illegible]-61.
DR P-PSDB; AAB44845.
XX
PT Novel 49 human secreted proteins useful for diagnosis, prevention and
PT treatment of disorders including neurological, cell proliferative,
PT cardiovascular, and autoimmune/inflammatory disorders and also for
PT infections
XX
PS Claim 1a; Page 343-344; 405pp; English.
XX
CC This invention describes a novel isolated polypeptide (I) comprising an
CC amino acid sequence at least 95 % identical to a polypeptide sequence
CC selected from 49 polypeptides encoded by polynucleotide sequences
CC included in American Type Culture Collection (ATCC) deposit number
CC 203917, defined in the specification. The products of the invention have
CC cytostatic, antiarthritic, antiasthmatic, immunosuppressive, neuroprotective,
CC antiatherosclerotic, antiinflammatory, neuroprotective, antidiabetic,
CC tranquiliser, vulnerary, antibacterial, antipsoriatic, antiarrhythmic,
CC antirheumatic, cardiant and anti HIV activity. (I) or a nucleic acid (II)
CC encoding (I) is useful for preventing, treating or ameliorating a medical
CC condition and for diagnosing a pathological condition or susceptibility
CC to the condition. (I) is useful for identifying a binding partner which
CC affects the activity of the polypeptide and for identifying an activity
CC in a biological sample. (i), (II) or an antibody (IV) specific for (I) is
CC also useful for treating or preventing a disease, disorder or condition
CC associated with aberrant expression of (I). Diseases treated or diagnosed
CC include immune disorders such as autoimmune diseases, blood protein
CC disorders, anemia, allergic reactions and conditions such as asthma,
CC organ rejection or graft-versus host disease, inflammation, hyper
CC proliferative disorders, cardiovascular disorders such as arteriovenous
CC fistula, arrhythmias, arteriosclerosis, coronary thrombosis, organ
CC regeneration, cancer, neovascular glaucoma, diabetic retinopathy,
CC rheumatoid arthritis, psoriasis, diseases associated with decreased
CC apoptosis that include acquired immunodeficiency syndrome (AIDS),
CC neurological diseases such as Parkinson's disease, viral, bacterial,
CC fungal or parasitic diseases. They are also used to repair, replace or
CC protect tissue damage by congenital defects, to treat trauma, as surgery,

[illegible] useful for purity, to treat illness, repetition injury, or systemic cytokine [illegible], to stimulate chondrocyte growth, to prevent skin injury due to sunburn, to change a mammal's mental state or physical state by antidepressant mechanisms, cardiac rhythms, depression, memory, stress, and for accelerating wound healing. (I), (II) and, or their agonist or antagonist are useful as food additives or preservatives to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamin, mineral or other nutritional components. (I) is useful for detecting therapeutic compounds. (II) is useful in forensic biology for detecting DNA sequences and as diagnostic probes for detecting the presence of specific mRNA in a particular cell type.

Sequence 1 of 80; 24 A; 7 C; 6 G; 5 T; 0 U; 0 Other;

Query Match 2.5%; Score 53; DB 1; Length 1;
Best Local Similarity 49.7%; Pred. No. 3.4;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 4 [illegible] 57

Db 4 [illegible] 403

Qy 58 [illegible] 111

Db 404 [illegible] 457

Qy 112 [illegible] 165

Db 458 [illegible] 511

RESULT 14
AAI61554/c
ID AAI61554 standard; DNA; 2561 BP.
XX
AC AAI61554;
XX
DT 06-NOV-1997 (first entry)
XX
DE Herpes virus DNA polymerase segment KSHV encoding DNA.
XX
KW Antiherpetic; food additive; herpes virus; detection; infection;
KW Kaposi's sarcoma herpes virus; viral DNA; viral RNA; vaccine;
KW antibody; mRNA.
XX
OS Kaposi's sarcoma herpes virus.
XX
FH Key Location/Qualifiers
FT misc_difference 1..2561
FT /note= "N = inosine"
FT misc_feature 1..2561
FT /product= "DNA polymerase"
FT /note= "N = inosine"
XX
SQ Sequence 2561 BP; 1024 A; 1024 C; 1024 G; 1024 T; 1 U; 1 Other;

Query Match 2.5%; Score 53; DB 1; Length 2561;
Best Local Similarity 49.7%; Pred. No. 3.4;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 4 [illegible] 57

Db 1258 [illegible] 1311

Qy 58 [illegible] 111

Db 1198 [illegible] 1251

Qy 112 [illegible] 165

Db 1138 [illegible] 1191

RESULT 15
AAV78802/c
ID AAV78802 standard; DNA; 45180 BP.
XX
AC AAV78802;
XX
DT 25-FEB-1999 (first entry)
XX
DE KSHV 160K DNA (nucleotides 1-45,180).
XX
KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; probe; PCR;
KW dihydrofolate reductase; DHFR; DNA polymerase; serine/threonine protein kinase;
KW diagnosis; treatment; HHV8; complement binding protein; gene 220;
KW ssDNA binding protein; transport protein; glycoprotein B; gpcB; 21.5;
KW DNA polymerase; viral interleukin 6; HHV4 IL6 1; thymidylate synthase;
KW vMIP II; HHV4 IEI II; vMIP I; capsid protein 1; tegument protein 1.8;
XX
OS Kaposi's sarcoma-associated herpesvirus.
XX
PN WO9945011-A.
XX
PD 10-SEP-1999.
XX
PF 29-NOV-1996; 96US-0770379.
XX
PR 29-NOV-1996; 96US-0770379.
XX
PA (UNIV-) UNIV COLUMBIA NEW YORK
XX
PI Boshoff PA, Chang Y, Ebelman IS, Moore PS, Boshoff PA;
XX
DR WPI; 1999-509711/06.
XX
PT Kaposi's sarcoma associated herpes virus nucleotide and [illegible]
PT dihydrofolate reductase and is useful for treatment, prophylaxis
PT or diagnosis of Kaposi's sarcoma

XX
PS Disclosure; Column 62 96; 109pp; English.
XX
CC This sequence is a fragment of the Kaposi's sarcoma associated
CC herpesvirus (KSHV) DHFR (long unique region). This fragment contains
CC coding regions for K1, ORF4 which encodes the complement binding protein
CC vCCBP, ORF5 which encodes a dsDNA binding protein (vSBP), ORF7 which
CC encodes a transport protein, ORF8 which encodes glycoprotein B, ORF9
CC which encodes DNA polymerase (pol), ORF10, ORF11, K2 which encodes viral
CC interleukin 6 (vIL-6), ORF62 which encodes dihydrofolate reductase
CC (DHFR), K3 which encodes BHV4-IE1 1, ORF70 which encodes thymidylate
CC synthase, K4 which encodes vMIP-II, K5 which encodes BHV4 IE1 11, R6
CC which encodes vMIP I, K7, vRF10 which encodes Bcl-2, ORF17 which encodes
CC capsid protein 1, ORF18 and ORF19 which encodes Tegument protein 1.
CC KSHV is a new human Herpesvirus (HHV8) believed to cause Kaposi's sarcoma
CC (KS) which is the most common form of neoplasm occurring in persons with
CC acquired immune deficiency syndrome (AIDS). The DHFR protein is useful
CC for vaccination, prophylaxis, diagnosis and treatment of a subject with
CC Kaposi's sarcoma and to detect [illegible] of a DNA virus associated
CC with Kaposi's sarcoma in a cell.
XX
SQ Sequence 35160 BP; 87,3 A; 9395 C; 8921 G; 8081 T; 0 other;

Query Match 2.5%; Score 33; DB 20; Length 35160;
Best Local Similarity 49.7%; Pred. No. 14;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 459 [illegible] 518
Db 13044 [illegible] 12975

Qy 519 [illegible] 578
Db 12974 [illegible] 12915

Qy 579 [illegible] 637
Db 12914 [illegible] 12866

Search completed: November 22, 2001, 16:08:22
Job time: 1.995 sec

GM protein - protein search, using sw model

Run on: November 23, 2001, 12:59:56 ; Search time 1241 Seconds
(without alignments)
5.218 Million cell updates/sec

Title:
Perfect score:
Sequence:

Scoring table:

Searched:

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length:

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: gb_est1_*
2: gb_est2_*
3: gb_est3_*
4: gb_est4_*
5: gb_est5_*
6: gb_est6_*
7: gb_est7_*
8: gb_est8_*
9: gb_est9_*
10: gb_est10_*
11: gb_est11_*
12: gb_est12_*
13: gb_est13_*
14: gb_est14_*
15: gb_est15_*
16: gb_est16_*
17: gb_est17_*
18: gb_est18_*
19: gb_est19_*
20: gb_est20_*
21: gb_est21_*
22: gb_est22_*
23: gb_est23_*
24: gb_est24_*
25: gb_est25_*
26: gb_est26_*
27: gb_est27_*
28: gb_est28_*
29: gb_est29_*
30: gb_est30_*
31: gb_est31_*
32: gb_est32_*
33: em_estba1_*
34: em_estba2_*
35: em_esthum1_*
36: em_esthum2_*
37: em_esthum3_*
38: em_esthum4_*
39: em_esthum5_*
40: em_esthum6_*
41: em_esthum7_*
42: em_esthum8_*
43: em_esthum9_*
44: em_esthum10_*
45: em_esthum11_*
46: em_esthum12_*
47: em_esthum13_*
48: em_esthum14_*
49: em_esthum15_*
50: em_esthum16_*
51: em_esthum17_*
52: em_esthum18_*
53: em_esthum19_*
54: em_esthum20_*
55: em_esthum21_*
56: em_esthum22_*
57: em_esthum23_*
58: em_esthum24_*
59: em_esthum25_*
60: em_esthum26_*
61: em_esthum27_*
62: em_esthum28_*
63: em_estinl_*
64: em_estin2_*
65: em_estin3_*
66: em_estin4_*
67: em_estin5_*
68: em_estoml_*
69: em_estom2_*
70: em_estovl_*
71: em_estov2_*
72: em_estpll_*
73: em_estpl2_*
74: em_estpl3_*
75: em_estpl4_*
76: em_estpl5_*
77: em_estpl6_*
78: em_estpl7_*
79: em_estpl8_*
80: em_estpl9_*
81: em_estpl10_*
82: em_estrol_*
83: em_estro2_*
84: em_estro3_*
85: em_estro4_*
86: em_estro5_*
87: em_estro6_*
88: em_estro7_*
89: em_estro8_*
90: em_estro9_*
91: em_estrol0_*
92: em_estrol1_*
93: em_estrol2_*
94: em_estrol3_*
95: em_estrol4_*
96: em_estrol5_*
97: em_estrol6_*
98: em_estrol7_*
99: em_estrol8_*
100: em_estrol9_*
101: em_estro20_*
102: gb_est25_*
103: gb_est26_*
104: gb_est27_*
105: gb_est28_*
106: gb_est29_*
107: gb_est30_*
108: gb_est31_*
109: gb_est32_*
110: gb_est41_*
111: gb_est42_*
112: gb_est43_*
113: gb_est44_*
114: gb_est45_*
115: gb_est46_*
116: gb_est47_*

117: qb_est49:*
118: qb_est50:*
119: qb_est51:*
120: qb_est52:*
121: qb_est53:*
122: qb_est54:*
123: qb_est55:*
124: qb_est56:*
125: qb_est57:*
126: qb_est58:*
127: qb_est59:*
128: qb_est60:*
129: qb_est61:*
130: qb_est62:*
131: qb_est63:*
132: qb_est64:*
133: qb_est65:*
134: qb_est66:*
135: qb_est75:*
136: qb_est76:*
137: qb_est77:*
138: qb_est78:*
139: qb_est79:*
140: qb_est80:*
141: qb_est81:*
142: qb_est82:*
143: qb_est83:*
144: qb_est84:*
145: qb_est85:*
146: qb_est86:*
147: qb_est87:*
148: qb_est88:*
149: qb_est89:*
150: qb_est90:*
151: qb_est91:*
152: qb_est99:*
153: qb_est100:*
154: qb_est101:*
155: qb_est102:*
156: qb_est103:*
157: qb_est104:*
158: qb_est105:*
159: qb_est106:*
160: qb_est67:*
161: qb_est68:*
162: qb_est69:*
163: qb_est70:*
164: qb_est71:*
165: qb_est72:*
166: qb_est73:*
167: qb_est74:*
168: qb_est92:*
169: qb_est93:*
170: qb_est94:*
171: qb_est95:*
172: qb_est96:*
173: qb_est97:*
174: qb_est98:*
175: qb_est108:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum41:*
179: em_esthum42:*
180: em_esthum43:*
181: em_estcm3:*
182: em_estpu11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_hte:*
187: qb_est107:*
188: qb_est108:*
189: qb_est109:*
190: qb_est110:*
191: qb_est111:*
192: qb_hte:*
193: em_qss_hum:*
194: em_qss_hum1:*
195: em_qss_hum2:*
196: em_qss_hum3:*
197: em_qss_hum4:*
198: em_qss_hum5:*
199: em_qss_hum6:*
200: em_qss_hum7:*
201: em_qss_hum8:*
202: em_qss_hum9:*
203: em_qss_inv1:*
204: em_qss_inv2:*
205: em_qss_inv3:*
206: em_qss_other:*
207: em_qss_plu1:*
208: em_qss_plu2:*
209: em_qss_pro:*
210: em_qss_rod1:*
211: em_qss_rod2:*
212: em_qss_rod3:*
213: em_qss_rod4:*
214: em_qss_rod5:*
215: em_qss_vri1:*
216: em_qss_vri2:*
217: em_qss_vri3:*
218: qb_qss1:*
219: qb_qss2:*
220: qb_qss3:*
221: qb_qss4:*
222: qb_qss5:*
223: qb_qss6:*
224: qb_qss7:*
225: qb_qss8:*
226: qb_qss9:*
227: qb_qss10:*
228: qb_qss11:*
229: qb_qss12:*
230: qb_qss13:*
231: qb_qss14:*
232: qb_qss15:*
233: qb_qss16:*
234: qb_qss17:*
235: qb_qss18:*
236: qb_qss19:*
237: qb_qss20:*
238: qb_qss21:*
239: qb_qss22:*
240: qb_qss23:*
241: qb_qss24:*
242: qb_qss25:*
243: qb_qss26:*
244: qb_qss27:*
245: qb_qss28:*
246: qb_qss29:*
247: qb_qss30:*
248: qb_qss31:*
249: qb_qss32:*
250: qb_qss33:*
251: qb_qss34:*
252: em_qss_inv4:*
253: em_qss_rod6:*
254: em_qss_rod7:*
255: em_qss_rod8:*
256: qb_qss35:*
257: qb_qss36:*
258: qb_qss37:*

Prob. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	888.2	29.2	888	106	BE520240	BE520240 MIIB12SIM
2	883	28.9	883	110	AW036132	AW036132 EST274508
3	883	28.9	883	110	AW036134	AW036134 EST274510
4	[illegible]	[illegible]	[illegible]	8	AA545747	AA545747 HBMSP1B1
5	[illegible]	[illegible]	[illegible]	[illegible]	BE436403	BE436403 EST407481
6	[illegible]	[illegible]	[illegible]	167	BE461872	BE461872 EST413291
7	[illegible]	[illegible]	[illegible]	153	BG457132	BG457132 NF100P05P
8	[illegible]	[illegible]	[illegible]	[illegible]	BE520241	BE520241 MIIB12XIM
9	[illegible]	[illegible]	[illegible]	135	BF177266	BF177266 EM1 1 702
10	[illegible]	[illegible]	[illegible]	[illegible]	H86613	H86613 yr44d12.s1
11	[illegible]	[illegible]	[illegible]	21	AI553154	AI553154 vi07f12.y
12	[illegible]	[illegible]	[illegible]	8	AA521521	AA521521 vi07b12.r
13	[illegible]	[illegible]	[illegible]	115	AW3[illegible]	AW3[illegible] MR3 SI019
14	[illegible]	[illegible]	[illegible]	[illegible]	AA[illegible]	AA[illegible] [illegible]
15	[illegible]	[illegible]	[illegible]	216	AZ[illegible]6678	AZ[illegible]6678 1M0410006
16	[illegible]	[illegible]	[illegible]	165	BE289814	BE289814 601089743
17	[illegible]	[illegible]	[illegible]	103	AI930492	AI930492 ul65e01.y
18	[illegible]	2.8	553	[illegible]	AW1[illegible]	AW1[illegible] [illegible]
19	[illegible]	2.8	618	228	AQ398287	AQ398287 mqxb00196
20	[illegible]	[illegible]	875	135	BF18[illegible]	BF18[illegible] 601809824
21	[illegible]	2.7	674	257	B93293	B93293 CIT HSP 217
22	[illegible]	2.7	701	156	BE536305	BE536305 602051954
23	[illegible]	2.7	[illegible]	150	BF581744	BF581744 602099525
24	[illegible]	2.7	[illegible]	8	AA479432	AA479432 zv14d11.r
25	[illegible]	2.7	238	118	AW[illegible]	AW[illegible] hn27h06.x
26	[illegible]	2.7	[illegible]	5	AA[illegible]	AA[illegible] EST21203
27	[illegible]	2.7	[illegible]	15	AA[illegible]772	AA[illegible]772 ad41h08.s
28	[illegible]	2.7	[illegible]	4	AA[illegible]772	AA[illegible]772 [illegible]
29	[illegible]	2.7	183	[illegible]	AW[illegible]	AW[illegible] MR3 SI020
30	[illegible]	2.7	[illegible]	[illegible]	AW3[illegible]	AW3[illegible] MR0 H1020
31	[illegible]	2.7	[illegible]	[illegible]	BE348769	BE348769 ht70d03.x
32	[illegible]	2.7	575	[illegible]	BE464[illegible]	BE464[illegible] [illegible]
33	[illegible]	2.7	[illegible]	155	BG547334	BG547334 602574753
34	[illegible]	2.7	[illegible]	173	BG110416	BG110416 602278825
35	[illegible]	2.7	1068	147	BF308318	BF308318 601890383
36	[illegible]	2.7	766	225	AQ162660	AQ162660 mqxb00133
37	85	2.6	172	115	AW377789	AW377789 MR0 H1020
38	85	2.6	580	23	AI692626	AI692626 wd86a12.x
39	84.4	2.6	159	1	AA044929	AA044929 zk77e11.r
40	84.4	2.6	423	138	BE669783	BE669783 7e35h04.x
41	84.4	2.6	425	136	BE503720	BE503720 7a31h06.x
42	84.4	2.6	505	24	AI761162	AI761162 wi58d03.x
43	84.4	2.6	507	105	AL133957	AL133957 DKFZp761
44	84.4	2.6	517	23	AI[illegible]	AI[illegible] wd8[illegible]04.x
45	84.4	2.6	525	9	AA603084	AA603084 np32h04.s

ALIGNMENTS

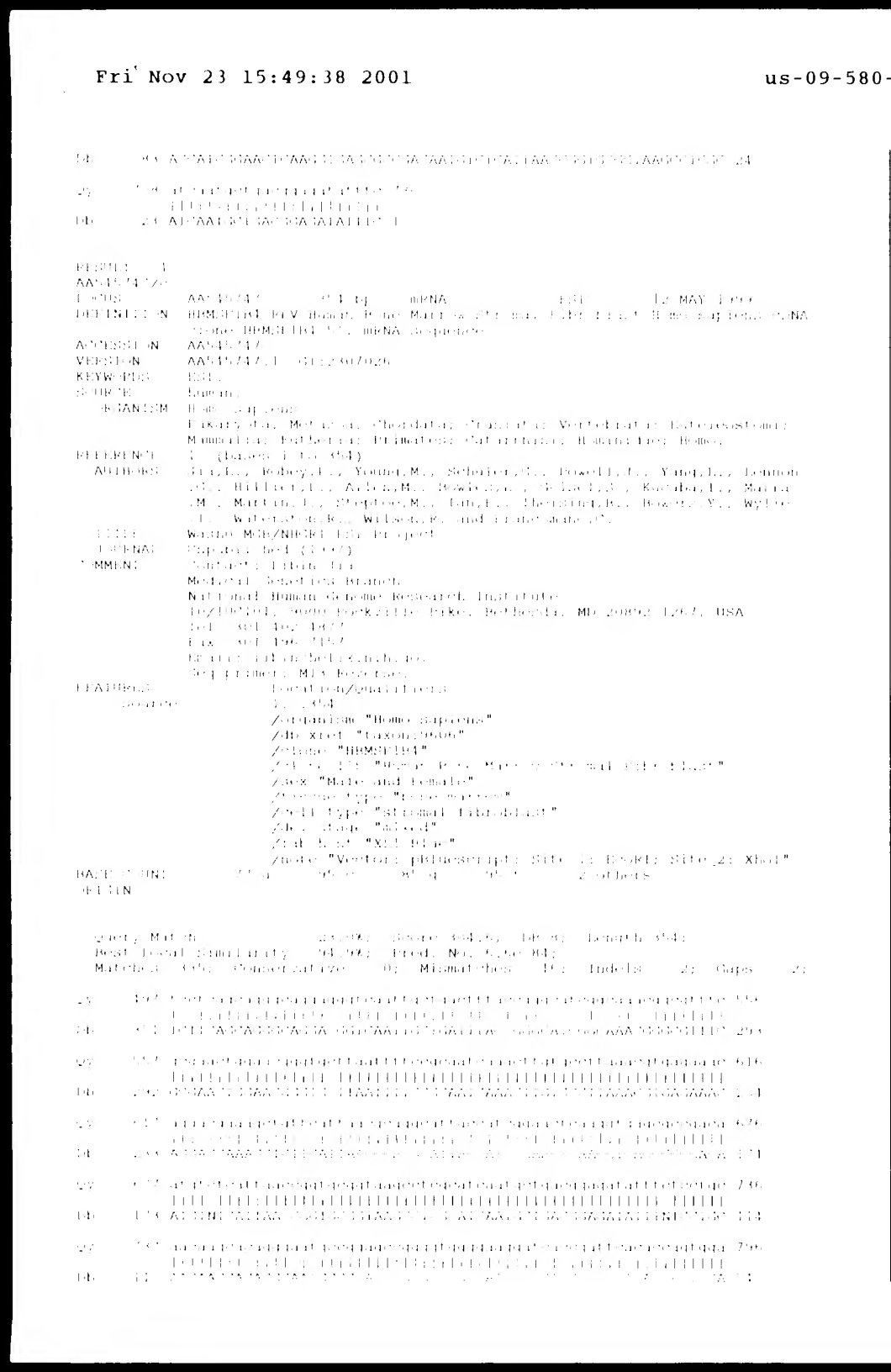
```
RESULT 1
BE520240/c
LOCUS       BE520240    [illegible] bp    mRNA            EST       [illegible]-MAR-2001
DEFINITION  MIIB12SIM Arabidopsis developing seed Arabidopsis thaliana cDNA
            clone MIIB12 5', mRNA sequence.
ACCESSION   BE520240
VERSION     BE520240.1  GI:[illegible]
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
  TITLE     A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed [illegible]
  JOURNAL   Plant Physiol. 124 (4), 1582-1594 (2000)
  MEDLINE   20567[illegible]
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI 48824
            , USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Michigan State University DNA Sequencing Facility Arabidopsis
            Biological Resource Center, The Ohio State University, [illegible] Botany &
            Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA  FAX:
            6142920[illegible] TEL: 614292937[illegible].
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism "Arabidopsis thaliana"
                     /strain "Columbia"
                     /db_xref "taxon:3702"
                     /clone "MIIB12"
                     /clone_lib "Arabidopsis developing seed"
                     /tissue_type "seed"
                     /dev_stage "5-13 days after flowering"
                     /lab_host "E.coli"
                     /note "Organ: Developing seeds Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoII"
BASE COUNT      [illegible] a    100 c    103 g    102 t    [illegible] others
ORIGIN

  Query Match             [illegible]%;  Score 888.2;  DB 106;  Length [illegible];
  Best Local Similarity   98.[illegible]%;  Pred. No. 2.3e-[illegible];
  Matches  888;  Conservative  [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy        [illegible] [illegible]
Db        [illegible] [illegible]

Qy        379 [illegible]
Db        333 [illegible]

Qy        439 [illegible] 498
Db        273 [illegible] 214

Qy        [illegible] [illegible] 558
Db        213 [illegible] 154

Qy        559 [illegible]
Db        153 [illegible] 94

Qy        619 [illegible]
Db        93 [illegible] 34

Qy        679 [illegible] 711
Db        33 [illegible] 1


RESULT 2
AW036132
LOCUS       AW036132    883 bp    mRNA            EST       [illegible]
DEFINITION  EST274508 tomato seed, TAMU Lycopersicon esculentum cDNA clone
            cLEE[illegible] similar to periplasmic phosphoanhydride phosphohydrolase
            precursor, putative, mRNA sequence.
ACCESSION   AW036132
VERSION     AW036132.1  GI:5894811
KEYWORDS    EST.
```

```
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 383)
  AUTHORS   Alcala,J., Vrebalov,J., White,K., Matern,A.I., Visser,I., Hall,I.H.
            , Liang,P., Upton,J., Bonnina,C.M., Traven,M.B., Pujii,C.Y., Bowman
            ,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
            ,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato seed tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5' prime sequence.
FEATURES             location/qualifiers
     source          1..383
                     /organism="lycopersicon esculentum"
                     /cultivar="IA496"
                     /db_xref="taxon:4081"
                     /clone="CLEEIE23"
                     /clone_lib="tomato seed, IAMF"
                     /tissue_type="seeds"
                     /dev_stage="quiescent seed"
                     /lab_host="XLI-Blue MRF'"
                     /note="Vector: pBluescript SK(-); Site 1: EcoRI Site 2:
                     XhoI; CLEE - Tomato Seed EST Library, Directionally cloned
                     cDNAs inserted into pBluescript SK(-) at 5' end with
                     EcoRI and 3' end with XhoI site."
BASE COUNT       102 a    131 c     97 g     53 t
ORIGIN
```

```
  Query Match             26.3%;  Score 488;  DB 11;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-108;
  Matches  383;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      348 taaaacaggcgaaacctttgccgccgaggctgggcaggcacctgactgtgcaataaccgtacatac 407
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 taaaacaggcgaaacctttgccgccgaggctgggcaggcacctgactgtgcaataaccgtacatac 60

Qy      408 ccaagccatatacatatccatccgatccgttattttaatcctctaaaaactgcatttaccca 467
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ccaagccatatacatatccatccgatccgttattttaatcctctaaaaactgcatttaccca 120

Qy      468 actgcatatacgcaaacatgcatcgatcgcgtatctcccaggcgaaacgaggtgcatgtcgca 527
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 actgcatatacgcaaacatgcatcgatcgcgtatctcccaggcgaaacgaggtgcatgtcgca 180

Qy      528 ctttaccaggcatcgaacgaaacgaacgtttccgaaactgaaacgaacgaatcttattttccgca 587
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 ctttaccaggcatcgaacgaaacgaacgtttccgaaactgaaacgaacgaatcttattttccgca 240

Qy      588 atcaaacttgtacttaaacgtaaaacgaaacgaaacgtgtcattaacgcgaggcatt 647
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 atcaaacttgtacttaaacgtaaaacgaaacgaaacgtgtcattaacgcgaggcatt 300

Qy      648 ac-atcgaaactcaaaatgaacgcaaatatcttattaacgaatgaaacgtcgc 702
            || |||| |||||||||||||||||||||||||||||||||||||||||||||
Db      301 acaatcgaaactcaaaatgaacgcaaatatcttattaacgaatgaaacgtcgc 355

Qy      703 atcaatgaaacgaaattttc 740
            ||||||||||||||||||||
Db      356 atcaatgaaacgaaattttc 378

RESULT  4
```

```
AW036134/c
LOCUS       AW036134        383 bp    mRNA            EST       17-SEP-1999
DEFINITION  EST274 Tomato Seed, IAMF Lycopersicon esculentum cDNA clone
            CLEEIE23 similar to peripheral phosphoantibiotic phosphatase type
            precursor, putative, mRNA sequence.
ACCESSION   AW036134
VERSION     AW036134.1  GI:6944813
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 383)
  AUTHORS   Alcala,J., Vrebalov,J., White,K., Matern,A.I., Visser,I., Hall,I.H.
            , Liang,P., Upton,J., Bonnina,C.M., Traven,M.B., Pujii,C.Y., Bowman
            ,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
            ,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato seed tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5' prime sequence.
FEATURES             location/qualifiers
     source          1..383
                     /organism="lycopersicon esculentum"
                     /cultivar="IA496"
                     /db_xref="taxon:4081"
                     /clone="CLEEIE23"
                     /clone_lib="tomato seed, IAMF"
                     /tissue_type="seeds"
                     /dev_stage="quiescent seed"
                     /lab_host="XLI-Blue MRF'"
                     /note="Vector: pBluescript SK(-); Site 1: EcoRI Site 2:
                     XhoI; CLEE - Tomato Seed EST Library, Directionally cloned
                     cDNAs inserted into pBluescript SK(-) at 5' end with
                     EcoRI and 3' end with XhoI site."
BASE COUNT       83 a     92 c    101 g    102 t
ORIGIN
```

```
  Query Match             26.3%;  Score 488;  DB 11;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-108;
  Matches  383;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      348 taaaacaggcgaaacctttgccgccgaggctgggcaggcacctgactgtgcaataaccgtacatac 407
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 taaaacaggcgaaacctttgccgccgaggctgggcaggcacctgactgtgcaataaccgtacatac 60

Qy      408 ccaagccatatacatatccatccgatccgttattttaatcctctaaaaactgcatttaccca 467
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ccaagccatatacatatccatccgatccgttattttaatcctctaaaaactgcatttaccca 120

Qy      468 actgcatatacgcaaacatgcatcgatcgcgtatctcccaggcgaaacgaggtgcatgtcgca 527
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 actgcatatacgcaaacatgcatcgatcgcgtatctcccaggcgaaacgaggtgcatgtcgca 180

Qy      528 ctttaccaggcatcgaacgaaacgaacgtttccgaaactgaaacgaacgaatcttattttccgca 587
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 ctttaccaggcatcgaacgaaacgaacgtttccgaaactgaaacgaacgaatcttattttccgca 240

Qy      588 atcaaacttgtacttaaacgtaaaacgaaacgaaacgtgtcattaacgcgaggcatt 647
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 atcaaacttgtacttaaacgtaaaacgaaacgaaacgtgtcattaacgcgaggcatt 300

Qy      648 ac-atcgaaactcaaaatgaacgcaaatatcttattaacgaatgaaacgtcgc 702
            || |||| |||||||||||||||||||||||||||||||||||||||||||||
Db      301 acaatcgaaactcaaaatgaacgcaaatatcttattaacgaatgaaacgtcgc 355

Qy      703 atcaatgaaacgaaattttc 740
            ||||||||||||||||||||
Db      356 atcaatgaaacgaaattttc 378

RESULT  5
```



```
Qy      478 [illegible] 533
Db        8 [illegible]

Qy      534 aaaa 537
            ||||
Db       [illegible] 17


RESULT      6
BE461872/c
LOCUS       BE461872       179 bp    mRNA            EST        27-JUL-2000
DEFINITION  EST413291 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
            clone cLEG40G22, mRNA sequence.
ACCESSION   BE461872
VERSION     BE461872.1  GI:[illegible]6174
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 179)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Bezzerides,I.E.,
            Liang,F., Hansen,I.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
            Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
            ,S.D.
  TITLE     Generation of ESTs from tomato fruit tissue, breaker stage
  JOURNAL   Unpublished (2000)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4906
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1..179
                     /organism="Lycopersicon esculentum"
                     /cultivar="Ailsa"
                     /db_xref="taxon:4081"
                     /clone="cLEG40G22"
                     /clone_lib="tomato breaker fruit, TIGR"
                     /tissue_type="Pericarp"
                     /dev_stage="breaker"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI;
                     Site_2: XhoI; Fruit were harvested at the breaker stage
                     (first sign of lycopene accumulation on the blossom end of
                     the fruit). Fruit were cut in half and the seeds and
                     locules were discarded prior to freezing the pericarp."
BASE COUNT       40 a     47 c     51 g     41 t
ORIGIN


  Query Match             [illegible];  Score [illegible];  DB [illegible];  Length 179;
  Best Local Similarity   [illegible];  Pred. No. [illegible];
  Matches  [illegible];  Conservative    0;  Mismatches    [illegible];  Indels    0;  Gaps    0;

Qy      398 ccgtacatacccaggaagatacgtccagtccgatccgttatttaatcctctaaaaactg 457
            ||||||||||||||||||||||||||||||||||||||||||||||.| |||||||||||||
Db      179 CCGTACATACCCAGGAAGATACGTCCAGTCCGATCCGTTATTTAATCCTCTAAAAACTG 120

Qy      458 [illegible] 517
Db      119 [illegible]

Qy      518 caattgctgacttacgggcatcggcaaacggcgtttcgcgaactggaacgggtgctt 576
            |||||||| |||| |||||||||||||||||||||||||||||||||||||||||||||
Db       59 [illegible]
```

```
RESULT      7
BG457132
LOCUS       BG457132       657 bp    mRNA            EST        16-MAR-2001
DEFINITION  NF100C05PL1F1036 Phosphate starved leaf Medicago truncatula cDNA
            clone NF100C05PL 5', mRNA sequence.
ACCESSION   BG457132
VERSION     BG457132.1  GI:13508457
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Liu,J., Scott,A.B., Harris,A.R., Gonzales,R.A., [illegible], Flores
            ,H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula phosphate starved leaf library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Harrison MJ
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7325
            Fax: 580 221 7380
            Email: mjharrison@noble.org
            Insert Length: 657   Std Error: 0.00
            Plate: 100  row: C  column: 05
            Seq primer: CACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1..657
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="NF100C05PL"
                     /clone_lib="Phosphate starved leaf"
                     /tissue_type="leaf"
                     /dev_stage="trifoliate"
                     /note="Vector: Lambda Zap; At the [illegible] M.
                     truncatula plants were transplanted to phosphate free sand
                     and grown for a further 30 days. During this [illegible]
                     period, the plants were fertilized [illegible] with 1/2
                     Hoaglands solution containing only 20uM potassium
                     phosphate. RNA was prepared from above ground tissues."
BASE COUNT      187 a    172 c    138 g    160 t
ORIGIN


  Query Match             11.9%;  Score 157.6;  DB 15;  Length 657;
  Best Local Similarity   97.2%;  Pred. No. 6.2e-38;
  Matches  171;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy      751 atgccgaagccggaatgggaaggatcaccgattcacaccagtggaacaccttgctaagt 810
            || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 [illegible] 540

Qy      811 [illegible] 869
Db      541 [illegible]

Qy      870 caccccgttattggatttgatcatggcagcgttgacgcccatccacccgcaaaaac 92[illegible]
            |||||||||||| |||||||||| | ||||||||||||||||||||||||||||||
Db      601 [illegible]


RESULT      8
BE520241
LOCUS       BE520241       [illegible] bp    mRNA            EST        [illegible]
DEFINITION  M11B12XIM Arabidopsis developing seed Arabidopsis thaliana cDNA
            clone M11B12 5', mRNA sequence.
ACCESSION   BE520241
VERSION     BE520241.1  GI:9778243
```

```
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 254)
  AUTHORS   White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
  TITLE     A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
  JOURNAL   Plant Physiol. 124 (4), 1582-1594 (2000)
  MEDLINE   20566808
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            204 Biochemistry, Michigan State University, East Lansing, MI 48824
              USA
            Tel: 517 353 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Michigan State University DNA Sequencing Facility Arabidopsis
            Biological Resource Center, The Ohio State University, 309 Botany &
            Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
            614.292.0636, TEL: 614.292.9471.
FEATURES             Location/qualifiers
     source          1..254
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="MIIH12"
                     /clone_lib="Arabidopsis developing seed"
                     /tissue_type="Seed"
                     /dev_stage="5-13 days after flowering"
                     /lab_host="E.coli"
                     /clone_lib="Arabidopsis developing seed; Vector: pBlue...ript SK ;
                     Site 1: EcoRI; Site 2: XhoII"
BASE COUNT       70 a     57 c     63 g     64 t
ORIGIN
```

```
  Query Match             10.8%;  Score 143.2;  DB 136;  Length 254;
  Best Local Similarity   94.9%;  Pred. No. 1.5e-33;
  Matches  148;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 atataaaagcatattaatcccattttatctttctgattcagttaaccccgcaatctgca 60
           ||.||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 ATGAAAGCGATCTTAATCCCATTTTATCTTTCTGATTCAGTTAACCCCGCAATCTGCA 129

Qy      61 ttcgctcagattgaagccgaagctgaagctgaaaattatgatgatgattgcagtcatcagat 120
           ||.|||||||||.|||||.|||||||||||||||||||||||||||||||||||||||||
Db     130 TTCGCTCAGATCGAAGCCGAAGCTGAAGCTGAAAATTATGATGATTGCAGTCATCAGAT 189

Qy     121 ctgatgatgctcaaccaggccagccaagctatgcaat 156
           || |||   | | | | | ||||||||||||||||
Db     190 CTGGTGATGCTGAACCAGGCCAGCCAAGCTATGCAAT 225


RESULT  9
BE772266
LOCUS       BE772266         473 bp    mRNA            EST       31-OCT-2000
DEFINITION  EST 3 Pco.11 A*2 Embryo 1 (EMI) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE772266
VERSION     BE772266.1  GI:11695133
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   Reid,S.E., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271 USA
            Tel: 706-542-1860
            Fax: 706-542-1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and reads
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN-KEV
            High quality sequence stop: 473
            PolyA No.
FEATURES             Location/qualifiers
     source          1..473
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EMI)"
                     /note="organ: Embryos germinated for 24 hr; Vector:
                     pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
                     EcoRI; The library was made from poly A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       125 a    122 c    118 g    108 t
ORIGIN
```

```
  Query Match             6.6%;  Score 87.8;  DB 115;  Length 473;
  Best Local Similarity   97.8%;  Pred. No. 3.5e-16;
  Matches   89;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    1210 tgtgaaggcgcaatgcgcaagcgcatgtatcagttgcgtgcagattttaagcgaaat 1269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGTGAAGGAGCGAATGCGCAAGCGCATGTATCAGTTGCGTGCAGATTTTAAGCGAAAT 60

Qy    1270 qaagcgcgcgataccgcgcgtgcgagttgcgat 1302
           ||||||||||||||||||||||||||||  ||
Db      61 GAAGCACGCATACCGGCGTGCAGTTGTGTAAT 91


RESULT  10
BG06613/c
LOCUS       BG06613          392 bp    mRNA            EST       06-OCT-1999
DEFINITION  yr43d12.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:208151 3' similar to gb:L14702|HUMSCALUE Human scRNA molecule
            , transcribed from Alu (tRNA); qb:J04970 CAKI-XYEPEPIDASE M
            PRECURSOR (HUMAN);contains Alu repetitive element; mRNA sequence.
ACCESSION   BG06613
VERSION     BG06613.1  GI:13134445
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 392)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lobos,S., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The Washu-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact:  Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 909
            High quality sequence stops: 401
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL; contact the
```

```
            IMAGE Consortium (http://image.llnl.gov) for further information.
            Insert Length: 989  Std Error: 0.00
            Seq primer: Primer1 -21m13
            High quality sequence stop: 301.
FEATURES             Location/Qualifiers
     source          1..392
                     /organism="Homo sapiens"
                     /db_xref="GDB:4777292"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:292161"
                     /clone_lib="Stanford human liver spleen (NEB)"
                     /sex="male"
                     /bio_mat_type="20 week post conception human"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site 1: Not I; Site 2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAATAATCTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adapters
                     (Pharmacia); digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization.  Library
                     constructed by Benito Soares and M.Fatima Botelho."
BASE COUNT       116 a    76 c    91 g    98 t      11 others
ORIGIN


  Query Match             5.1%;  Score 47.6;  DB 21;  Length 992;
  Best Local Similarity   51.7%;  Pred. No. 0.049;
  Matches   93;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy      519 aattgctaatttttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 578
            | ||  ||  |||||  |||||  |||||||||||||||||||||||||||||||||||||
Db      252 AAGGTTTGGGGGGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 311

Qy      579 ttttccgcgaatgcaaaacttatgccttaaacgtaaaaaaaaaaaaaaaaaaaaaaaaaaa 638
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      202 TTATGGCGCTTACGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 143

Qy      639 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 698
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      142 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 83


RESULT 11
AI553154/c
LOCUS       AI553154      431 bp    mRNA            EST       23-MAR-1999
DEFINITION  oi97e12.y1 Pooled mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1063871 5' similar to gb:M90356 MYOSIN HEAVY CHAIN MOUSE (HUMAN),
            mRNA sequence.
ACCESSION   AI553154
VERSION     AI553154.1  GI:3119***
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Peek,K., Wylie,T.,
            Underwood,K., Strong,M., Theising,B., Allen,M., Bowers,Y., Porson
            ,B., Swaller,L., Gibbons,M., Pape,D., Harvey,N., Schalk,K., Ritter
            ,E., Kohn,S., Shinn,L., Jackson,Y., Cardenas,M., McCann,P.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/Waterston RH Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            M13-62x735
            This read is a RESEQUENCE of a previously sequenced Stanford
            This read has been verified (found to hit its original spot in the
            correct orientation)
            Seq primer:  -40RP from Gibco
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1..431
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1063871"
                     /clone_lib="Stanford mouse myotubes MPLRB5"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAATAATCTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adapters
                     [AATTCGGCACGAG-OH], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead.  The C2C12 cell line
                     (available from ATCC, catalog # CRL 1772) differentiates
                     rapidly, forming contractile myotubes and producing
                     characteristic muscle proteins.  "
BASE COUNT        76 a   146 c   126 g    82 t      1 others
ORIGIN


  Query Match             4.9%;  Score 45.8;  DB 21;  Length 431;
  Best Local Similarity   53.2%;  Pred. No. 0.34;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy      126 ctttaaaccaaaggcaacgcaaatgatgcaaggatgtgaccccaaatgcgatgcgcgatgc 185
            |||| ||  |||| ||| | ||| | |||| || ||||| | |||| ||| || |||
Db      366 CTGCCTCCCCTGGCCTTGCTAGTGAGGAAGGAGGTGTCCCCAAAGGCGTCCCCAGCCCCGT 307

Qy      186 cgatgaaagtgagtgagtgagtaacaccgcaaagtgatgaaactaatgcgctatgcgtttgatt 245
            |  |  ||| |||| ||  ||  | ||| | || ||| ||  |  |  |  |  |  |
Db      306 GCAGCAGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 247

Qy      246 ggcaacgcgcaagctcgcgtcaacgcaaatgatgcgtga 283
            |  |||  |||| |  | |   | |  ||||
Db      246 GCGCAGGTGAGCGGCGTGAGTCAGAGAGG 217

RESULT 12
AA521521/c
ID   AA521521      600 bp    mRNA            EST       17-JUL-1997
DEFINITION  zk71e12.r1 Soares mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1063871 5', similar to gb:M90356 MYOSIN HEAVY CHAIN MOUSE (HUMAN),
            mRNA sequence.
ACCESSION   AA521521
VERSION     AA521521.1  GI:2262266
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Hillier,N., Kucaba,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Meyer,R.,
            Schellenberg,K., Stapleton,M., Tan,F., Underwood,K., Porson,B.,
            Theising,B., Wylie,T., Yennamalla,S., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
```

```
            Fax: 314 286 0810
            Email: mcombs@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            NCI_CGAP
            Seq primer: pUC18 rev, EI from Amersham
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1..2007
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:[illegible]"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="Brain; glioblastoma"
                     /clone_lib="NCI_CGAP_Brn25"
                     /dev_stage="Adult"
                     /cell_line="CCF12"
                     /sex="Male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac (Pharmacia) with a modified
                     polylinker; Site 1: BssHII; Site 2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer (5'
                     [illegible]GGAAG)30VN-3'); double-stranded cDNA was ligated to Bss HII adaptors
                     (AATTCCGAGCGG), digested with Not I and cloned into the
                     Not I and Bss HII sites of the modified pT7T3 vector.
                     Library constructed by Bento Soares. The CCF12 cell line
                     (available from ATCC, catalog # CRL 1772) differentiates
                     rapidly, forming neurites and processes and producing
                     characteristic muscle proteins. "
BASE COUNT       [illegible] a    [illegible] c    [illegible] g    [illegible] t
ORIGIN


  Query Match                 [illegible]%;  Score 1[illegible]; DB 11;  Length [illegible];
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches  [illegible]; Conservative    0; Mismatches   [illegible]; Indels    [illegible]; Gaps    0;

Qy      [illegible] [illegible]
            [illegible]
Db      [illegible] [illegible]

Qy      [illegible] [illegible]
            [illegible]
Db      [illegible] [illegible]

Qy      [illegible] [illegible]
            [illegible]
Db      [illegible] [illegible]


RESULT [illegible]
AW[illegible]/c
LOCUS       AW[illegible]      [illegible] bp    mRNA            EST       04-FEB-2000
DEFINITION  MR[illegible] [illegible] [illegible] Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW[illegible]
VERSION     AW[illegible].1  GI:[illegible]
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   [illegible]
  TITLE     The [illegible]/LLNL Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.; Matthew
            Bowen, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            [illegible]
            Insert Length: 498  Std Error: 0.00
            Seq primer: 40mi8 fwd, EI from Amersham.
FEATURES             Location/Qualifiers
     source          1..498
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:[illegible]"
            [illegible]
            Fax: [illegible]
            Email: [illegible]
            This sequence was identified from the LLNL Human Cancer Genome
            Project. This entry can be seen in the following URL:
            [illegible]
            Seq primer: puc 18 forward
            High quality sequence start: 18
            High quality sequence stop: 360.
FEATURES             Location/Qualifiers
     source          1..504
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="SCOT2"
                     /dev_stage="Adult"
                     /note="Organ: stomach; Vector: pT7T3-Pac (Pharmacia) with a
                     Site 1: BssHII; Site 2: SmaI; A mini library was made to create a set of
                     derived from CRESTED PCR (U.S. letters latest application
                     No. 19[illegible]  Ludwig Institute for Cancer Research)
                     profiles into the pBC 18 vector. Reverse transcriptase
                     Tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       [illegible] a    [illegible] c    [illegible] g    [illegible] t    [illegible] others
ORIGIN


  Query Match                 [illegible]%;  Score [illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches  [illegible]; Conservative    0; Mismatches   [illegible]; Indels    [illegible]; Gaps    0;

Qy      [illegible] [illegible]
            [illegible]
Db      [illegible] [illegible]

Qy      [illegible] [illegible]
            [illegible]
Db      [illegible] [illegible]

Qy      [illegible] [illegible]
            [illegible]
Db      [illegible] [illegible]


RESULT 14
AA915940/c
LOCUS       AA915940           398 bp    mRNA            EST       [illegible]
DEFINITION  oa14d06.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:[illegible] similar to
            mRNA sequence.
ACCESSION   AA915940
VERSION     AA915940.1  GI:[illegible]
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.; Matthew
            Bowen, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            [illegible]
            Insert Length: 498  Std Error: 0.00
            Seq primer: 40mi8 fwd, EI from Amersham.
FEATURES             Location/Qualifiers
     source          1..398
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:[illegible]"
```

```
                     /clone_lib "NCI_CGAP_[illegible]"
                     /[illegible] "[illegible]"
                     /lab_host "DH10B"
                     /note "[illegible] Vector: pT7T3D [illegible] (Pharmacia) with a
                     modified polylinker; [illegible]
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. [illegible]
                     to Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library is normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT     [illegible] a   12[illegible] c   123 g   92 t
ORIGIN

  Query Match             2.9%;  Score 38.[illegible];  DB 13;  Length 398;
  Best Local Similarity   54.6%;  Pred. No. [illegible].78;
  Matches   77;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        [illegible]5 [illegible] 721
Db        390 [illegible]AAAAAAA[illegible] 331

Qy        725 tattttctcctgcaacaagcacagggaatgccggagccgggtggggaaggatcaccgatt 784
Db        [illegible] 27[illegible]

Qy        785 [illegible] 8[illegible]
Db        2[illegible] 25[illegible]


RESULT 1[illegible]
AZ[illegible]678/c
LOCUS       AZ[illegible]
DEFINITION  [illegible] Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M 41[illegible], DNA sequence.
ACCESSION   AZ[illegible]678
VERSION     AZ[illegible]
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,[illegible], Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pederson,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (200[illegible])
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 [illegible]
            Fax: 801 585 7177
            Email: [illegible]genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: [illegible]  row: [illegible]  column: [illegible]
            Seq primer: CGTTGTAAAACGACGGCCAG[illegible]
            Class: plasmid ends
            High quality sequence stop: [illegible].
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism "Mus musculus"
                     /strain "C57BL/6J"
                     /db_xref="taxon:[illegible]"
                     /clone "UUGC1M041[illegible]"
                     /clone_lib "Mouse 10kb plasmid UUGC1M library"
                     /sex "Male"
                     /lab_host "E. Coli strain XL10 Gold, [illegible]"
                     /note "Vector: [illegible]; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     ([illegible]). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     [illegible]. The sheared DNA
                     was blunt end repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a [illegible] to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically competent E. coli XL10 Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT     134 a   211 c   177 g   147 t
ORIGIN

  Query Match             2.9%;  Score 38.4;  DB 246;  Length [illegible];
  Best Local Similarity   4[illegible];  Pred. No. 1;
  Matches  102;  Conservative    0;  Mismatches  106;  Indels    [illegible];  Gaps    0;

Qy        253 cgccagcgtctggtagccgacggattgctggcgaaaaagggctgcccgcagtctggt[illegible] [illegible]
Db        569 CGGA[illegible] [illegible]

Qy        313 gtcgcgattattgctgatgtcgacgacgtaccgtaaacaggcgaagcctt[illegible] [illegible]2
Db        [illegible] 45[illegible]

Qy        373 ggg[illegible]ctgactgtgcaataaccgtacatacccaggcaggtacgtccagtcc[illegible] 4[illegible]2
Db        44[illegible] 3[illegible]

Qy        433 ccgttatttaatcctctaaaactgacg 460
Db        3[illegible] 362


Search completed: November 22, 2001, 15:32:49
Job time: [illegible] sec
```

GeneCore version 1.5
Copyright (c) 1993 - 2000 Compugen Ltd.

SW protein - protein search, using sw model

Run on: November 20, 2001, 09:59:23 ; Search time 2.50 Seconds
(without alignments)
1652.440 Million cell updates/sec

Title:
Perfect score:
Sequence: 1 MKAILIFPEESLEIPLIPQSA ... ... IVNPAAIPAVCIESHHHHHHH 440

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219441 seqs, 76124952 residues

Total number of hits satisfying chosen parameters: 219441

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

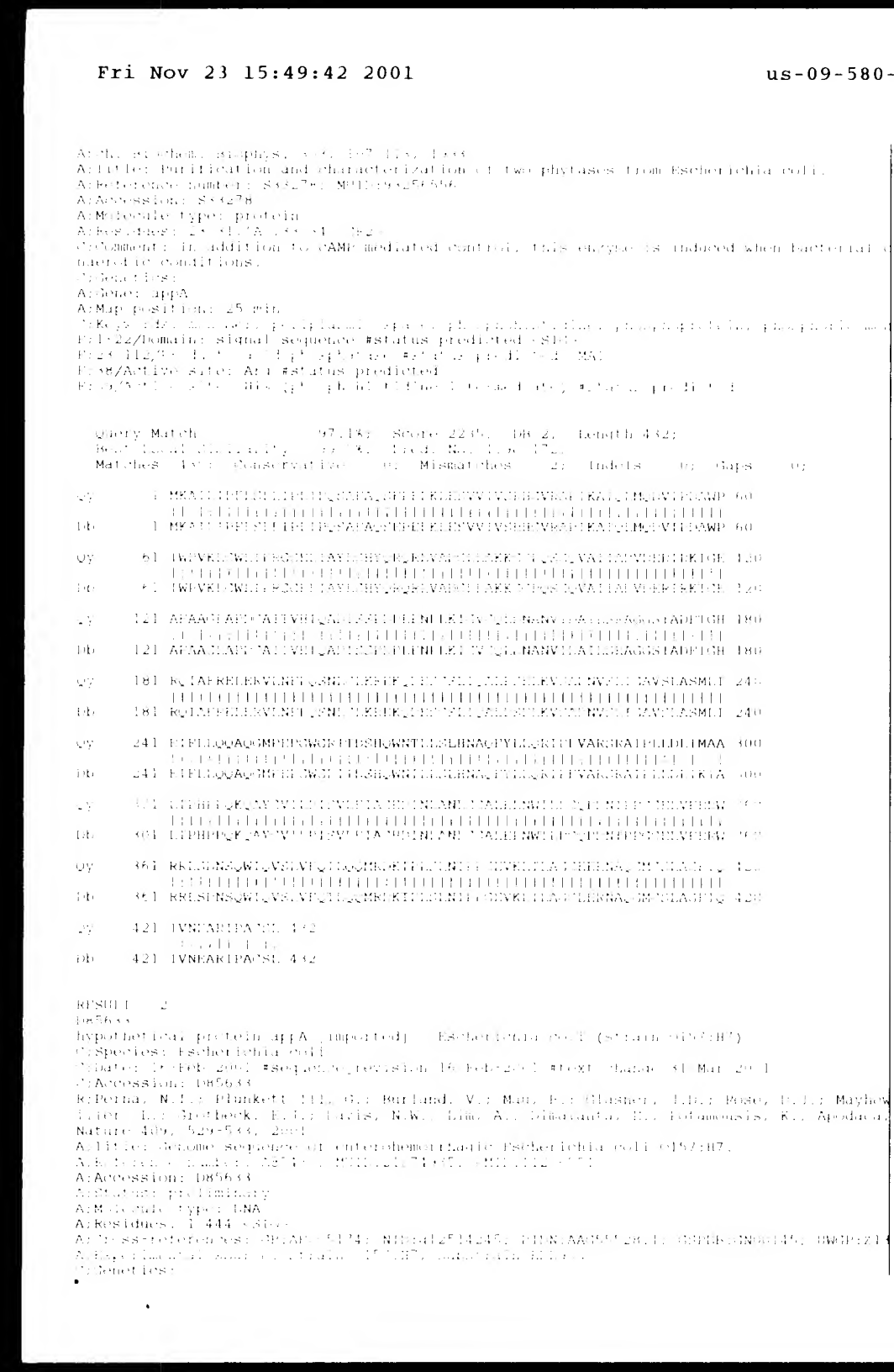
Prot N... is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is calculated by analysis of the total score distribution.

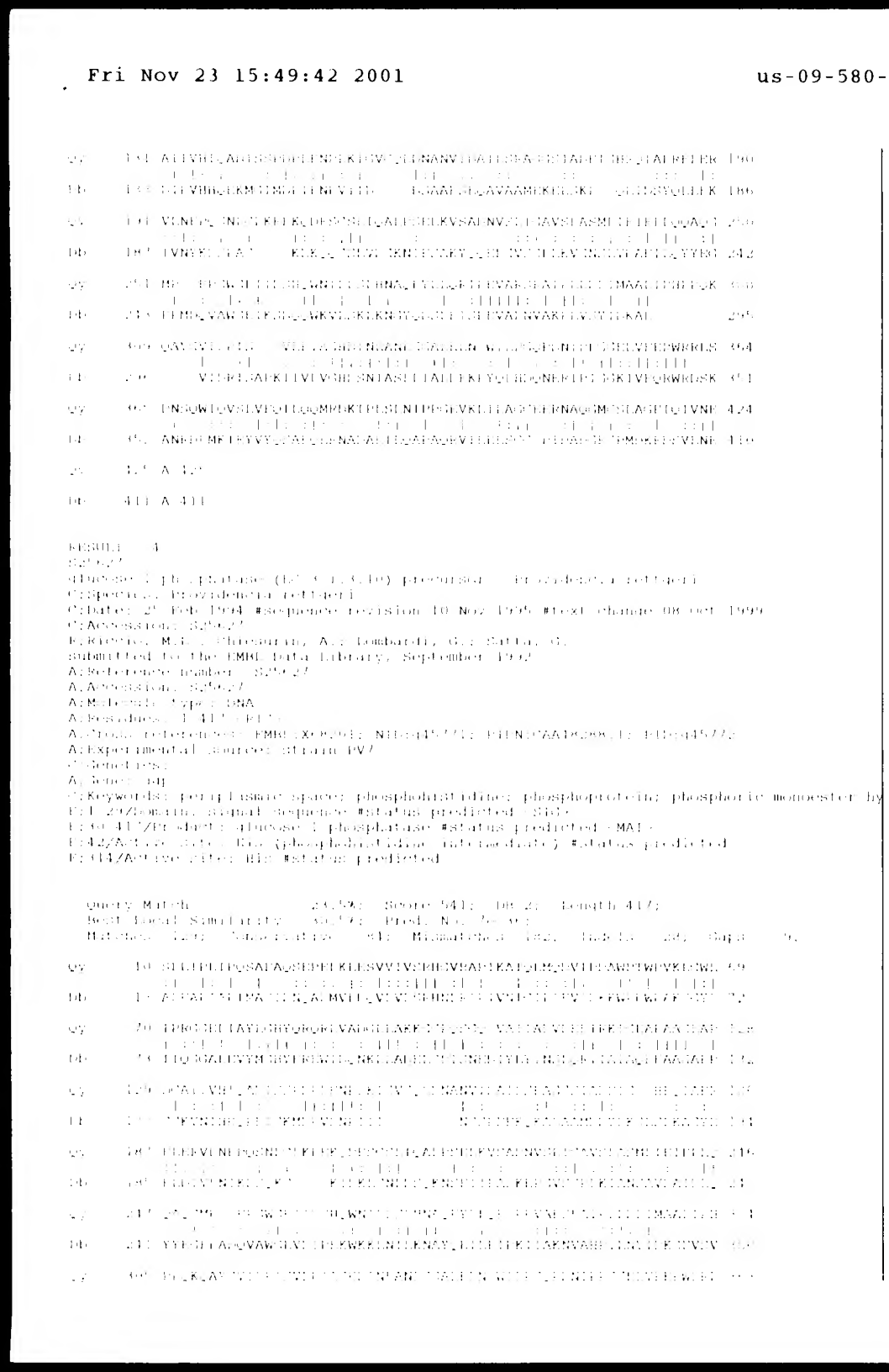
SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2235	97.1	432	2	B36733	acid phosphatase (
2	2210	96.0	414	2	D85633	hypothetical prote
3	545	23.7	314	2	B85536	periplasmic glucos
4	541	23.5	417	2	S25627	glucose-1-phosphat
5	509	22.1	418	2	JV0987	glucose-1-phosphat
6	155.5	6.8	4.3	2	S04167	acid phosphatase (
7	133	5.8	416	2	T10658	hypothetical prote
8	132	5.7	4.3	2	A43395	acid phosphatase (
9	122.5	5.3	4.1	2	S14742	acid phosphatase (
10	117	5.1	386	1	JH0610	acid phosphatase (
11	109	4.7	2554	1	VVIFPL	kinase related pro
12	108.5	4.7	10228	2	I30225	polyketide synthas
13	107.5	4.7	457	2	A56925	paired box transcr
14	107	4.6	450	2	A54429	paired box transcr
15	106.5	4.6	1048	1	BVBC5C	cytochrome (B7 s.
16	106.5	4.6	459	2	S52250	paired box transcr
17	102	4.4	698	2	A29440	signal recognition
18	101.5	4.4	367	2	C81416	hypothetical prote
19	101.5	4.4	479	1	JN0890	acid phosphatase (
20	101.5	4.4	521	2	H75536	probable acid--CoA
21	101.5	4.4	1047	2	S65245	ATP-dependent dsDN
22	101	4.4	5662	2	T18540	motA protein precu
23	101	4.4	2095.*	1	I38414	fatty-acid-ami mus
24	100.5	4.4	909	2	JQ0228	hydroxymethylbilan
25	100.5	4.4	438	2	S04682	acid phosphatase (
26	100.5	4.4	438	2	S08415	major plasmid tran
27	100.5	4.4	6260	2	I30228	polyketide synthas
28	100	4.4	498	2	A55752	ribosomal-protein-
29	100	4.4	1374	1	VCBE17	major capsid proto
30	100	4.4	4145	2	I42629	telomeric X- locus
31	99	4.3	646	2	G01983	pilin-like protein
32	98.5	4.3	425	2	D84306	hypothetical prote
33	97.5	4.2	578	2	H75256	malate oxidoreduct
34	97	4.2	449	2	I32457	hypothetical prote
35	97	4.2	587	2	B65171	hypothetical prote
36	97	4.2	1695	2	I10823	hypothetical prote
37	96.5	4.2	981	2	JH0152	acid-CoA ligase (
38	96.5	4.2	2856	2	A48654	probable polyketid
39	96.5	4.2	2041	2	I17439	peptide synthetase
40	96	4.2	942	2	S75598	poly(A) polymerase
41	96	4.2	4006	2	I00020	probable transmemb
42	95.5	4.1	234	2	F42696	thioredoxin (EC 1.8
43	95.5	4.1	356	2	F75564	hypothetical protein
44	95.5	4.1	479	1	JN0715	4-phytase (EC 3.1.
45	95.5	4.1	741	2	A84221	hypothetical prote

ALIGNMENTS

RESULT 1
B36733
acid phosphatase (EC 3.1.3.2) precursor - Escherichia coli
N;Alternate names: phosphoanhydride phosphohydrolase; periplasmic phosphatase
C;Species: Escherichia coli
C;Date: 19-Apr-1991 #sequence_revision 19-Apr-1991 #text_change 08-Jul-1999
C;Accession: B36733; S18018; B64809; A26534; S17956; S33278
R;Ishiwa, J.; Marek, C.; Bosquet, P.L.
J. Bacteriol. 172, 5497-5500, 1990
A;Title: The complete nucleotide sequence of the Escherichia coli gpsA gene: gpsA is co-s
A;Reference number: A36733; MUID:90386616
A;Accession: B36733
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-432 <DAS>
A;Cross-references: GB:M58708; NID:g145283; PIDN:AAA72009.1; PID:g145284
R;Greiner, R.; Jany, K.D.
Biol. Chem. Hoppe-Seyler 372, 664-665, 1991
A;Title: Characterization of a phytase from Escherichia coli
A;Reference number: S18018
A;Accession: S18018
A;Molecule type: protein
A;Residues: 23-33 <GRE>
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12
A;Reference number: A64720; MUID:97426017
A;Accession: B64809
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-432 <BLA>
A;Cross-references: GB:AE000208; GB:U00096; NID:g2367111; PIDN:AAC73002.1; PID:g2367
A;Experimental source: strain K-12, substrain MG1655
R;Touati, E.; Danchin, A.
Biochimie 69, 215-221, 1987
A;Title: The structure of the promoter and amino-terminal region of the pE-5 acid p
A;Reference number: A26534; MUID:87221766
A;Accession: A26534
A;Molecule type: DNA
A;Residues: 1-50,'NAGCHFKRMANLAG',65,'I',67-74,'DV',77-111,'S'-440
A;Cross-references: GB:X59471; NID:g48992; PIDN:CAA33431.1; PID:g48
R;Iwasa, J.; Ezaki, H.; Marek, C.; Dion, M.; Kletter Bontemps, M.; Bopet, P.
Mol. Gen. Genet. 229, 441-452, 1991
A;Title: A new oxygen-regulated operon in Escherichia coli comprises two genes f
A;Reference number: S17956; MUID:92049231
A;Accession: S17956
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-17 <DAZ>
A;Cross-references: GB:X76431; NID:g295256; PIDN:CAA53911.1; PID:g295
R;Greiner, R.; Konietzny, U.; Jany, K.D.



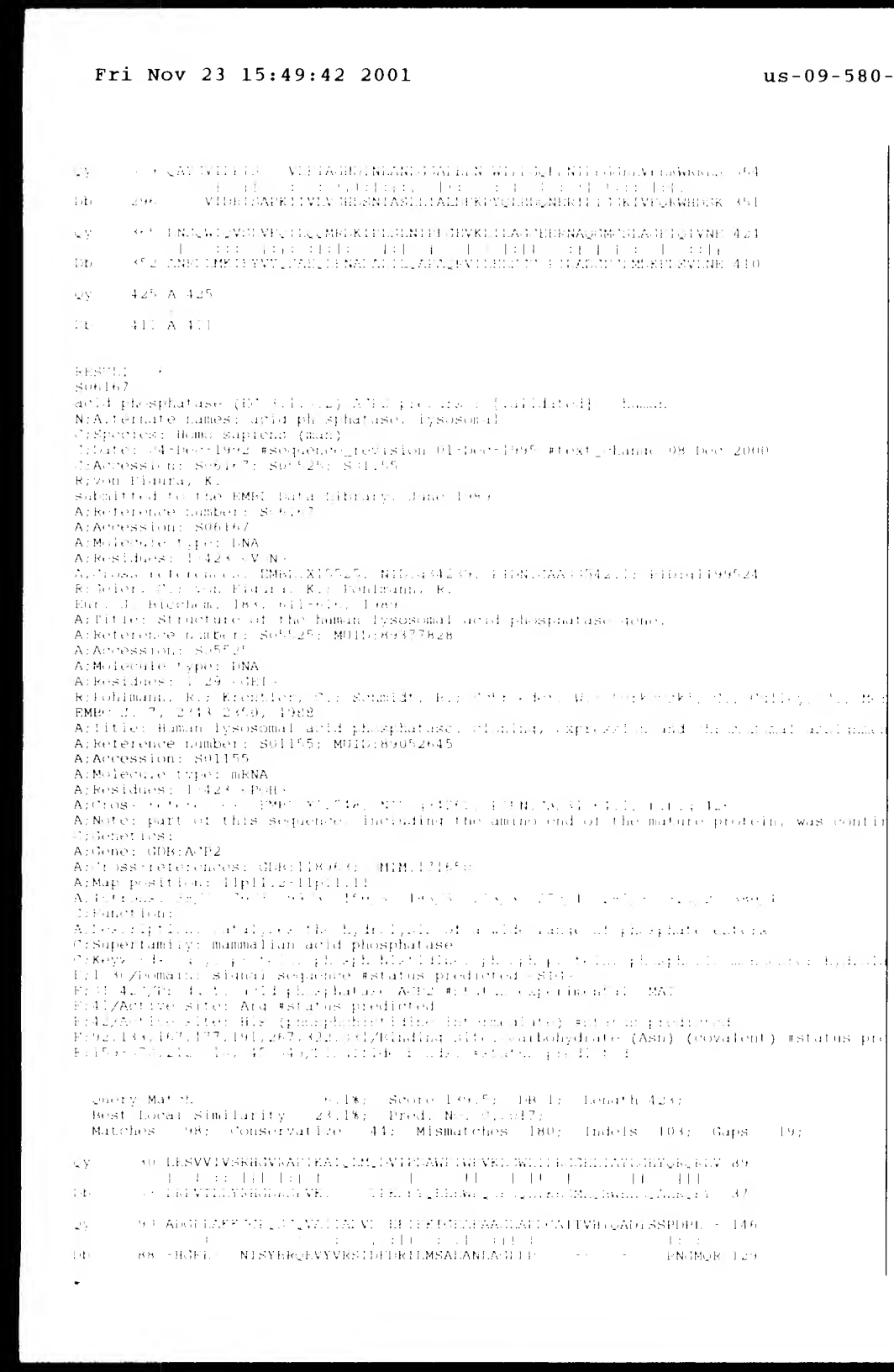


RESULT 4
S25627
alucose 1-phosphatase (EC 3.1.3.10) precursor - Escherichia coli
C;Species: Escherichia coli
C;Date: 27 Feb 1994 #sequence_revision 10 Nov 1995 #text_change 08 Oct 1999
C;Accession: S25627
R;Rieseo, M.E.; Chiesurin, A.; Tombaroli, G.; Satta, G.
Submitted to the EMBL Data Library, September 1992
A;Reference number: S25627
A;Accession: S25627
A;Molecule type: DNA
A;Residues: 1-417 <RIE>
A;Cross-references: EMBL:X60260; NID:g45771; PIDN:CAA42888.1; PID:g45772
A;Experimental source: strain PV7
C;Genetics:
A;Gene: agp
C;Keywords: periplasmic space; phosphohistidine; phosphoprotein; phosphoric monoester hy
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-417/Product: glucose 1-phosphatase #status predicted <MAT>
F;240/Active site: His (phosphohistidine intermediate) #status predicted
F;411/Active site: His #status predicted

Query Match 23.5%; Score 541; DB 2; Length 417;
Best Local Similarity 42.2%; Pred. No. 7e-45;
Matches 125; Conservative 75; Mismatches 161; Indels 29; Gaps 6;

RESULT 5
JV0087
glucose 1-phosphatase (EC 3.1.3.10) precursor - Escherichia coli
C;Species: Escherichia coli
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-1999
C;Accession: JV0087; BF1843
R;Brzeski, H.; Marek, C.; Bosqet, P.L.
J. Bacteriol. 172, 802-807, 1990
A;Title: Nucleotide sequence and transcriptional analysis of the
A;Reference number: JV0087; MUID:90140318
A;Accession: JV0087
A;Molecule type: DNA
A;Residues: 1-413 <BRA>
A;Cross-references: GB:M33807; NID:g145217; PIDN:AAA24126.1; PID:g145218
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12
A;Reference number: A64720; MUID:97426617
A;Accession: B64841
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-413 <BLA>
A;Cross-references: GB:AE000282; GB:U00096; NID:g1787233; PIDN:AAC74021.1; PID:g1787
A;Experimental source: strain K-12; substrain MG1655
C;Genetics:
A;Gene: agp
A;Map position: 23 min
C;Function:
A;Description: essential for growth in a high phosphate medium containing glucose 1-p
A;Note: optimal at low pH
C;Keywords: homodimer; periplasmic space; phosphohistidine; phosphoprotein; phosphori
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-413/Product: glucose 1-phosphatase #status predicted <MAT>
F;240/Active site: His (phosphohistidine intermediate) #status predicted
F;411/Active site: His #status predicted

Query Match 23.4%; Score 539; DB 2; Length 413;
Best Local Similarity 42.1%; Pred. No. 1e-45;
Matches 125; Conservative 75; Mismatches 161; Indels 29; Gaps 6;

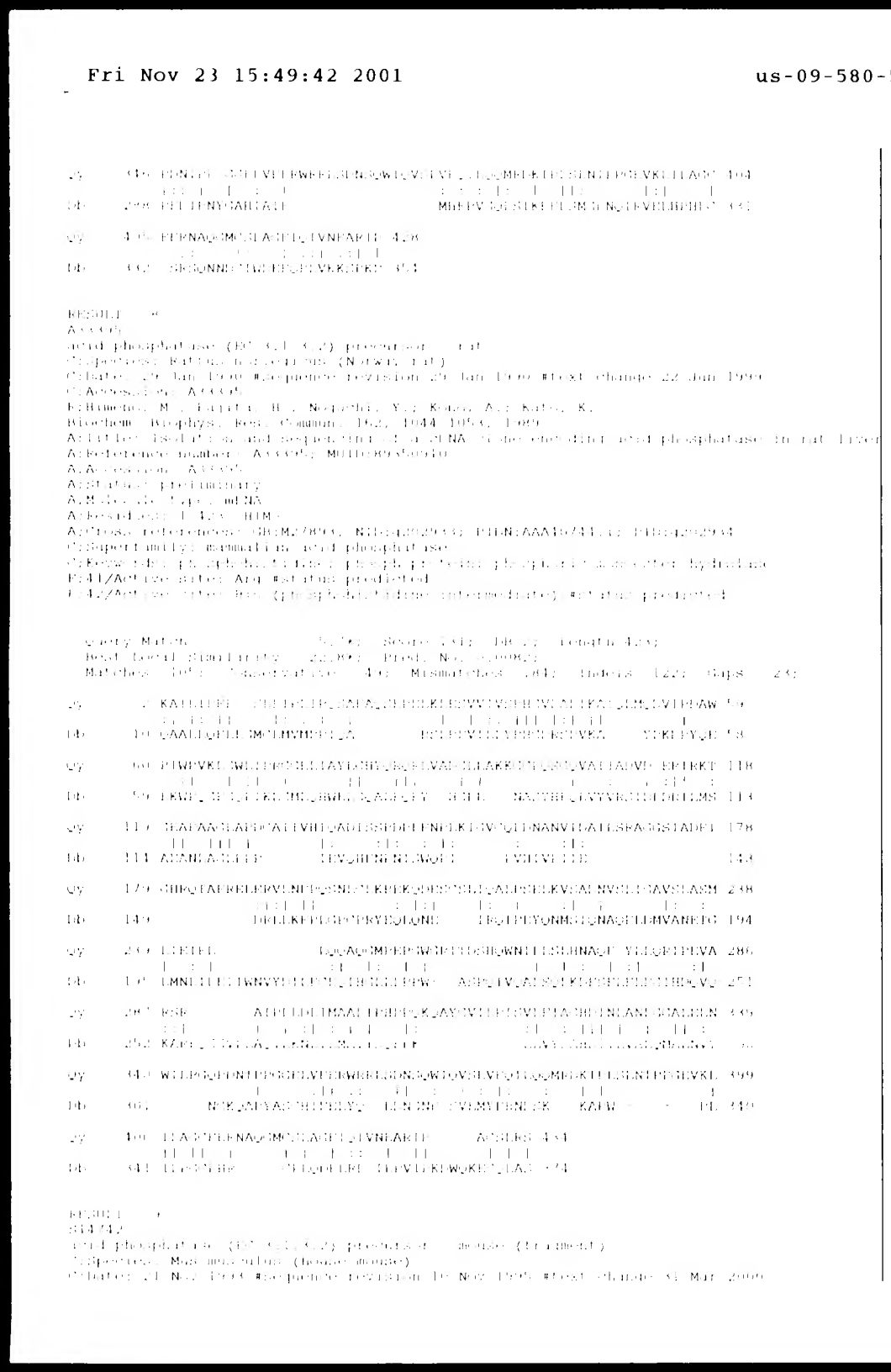


RESULT 1
S06167
acid phosphatase (EC 3.1.-.-) A precursor - human (fragment)
N;Alternate names: acid phosphatase, lysosomal
C;Species: Homo sapiens (man)
C;Date: 04-Dec-1992 #sequence_revision 01-Dec-1995 #text_change 08-Dec-2000
C;Accession: S06167; S05525; S11155
R;von Figura, K.
Submitted to the EMBL Data Library, June 1989
A;Reference number: S06167
A;Accession: S06167
A;Molecule type: DNA
A;Residues: 1-423 <VON>
A;Cross-reference: EMBL:X15256; NID:g42476; PIDN:CAA33412.1; PID:g41999
R;Seiler, J.; von Figura, K.; Hohlbaum, K.
Eur. J. Biochem. 188, 611-616, 1990
A;Title: Structure of the human lysosomal acid phosphatase gene.
A;Reference number: S05525; MUID:90377828
A;Accession: S05525
A;Molecule type: DNA
A;Residues: 1-29 <SEI>
R;Lohmann, K.; Krentler, C.; Schmidt, B.; [illegible]; [illegible]; [illegible]
EMBO J. 7, 2743-2750, 1988
A;Title: Human lysosomal acid phosphatase, cloning, expression and localization
A;Reference number: S01155; MUID:89062645
A;Accession: S01155
A;Molecule type: mRNA
A;Residues: 1-423 <POH>
A;Cross-reference: EMBL:X15256; NID:g42476; PIDN:CAA33412.1; PID:g41999
A;Note: part of this sequence, including the amino end of the mature protein, was contin
C;Genetics:
A;Gene: GDB:ACP2
A;Cross-reference: GDB:119064; MIM:171650
A;Map position: 11p11.2-11p15.1
A;[illegible]
C;Function:
A;Description: catalyzes the hydrolysis of a wide range of phosphate esters
C;Superfamily: mammalian acid phosphatase
C;Keywords: [illegible]
F;1-9/Domain: signal sequence #status predicted <SIG>
F;1-423/Gene: GDB:ACP2 #status experimental <EXP>
F;1-423/Active site: Asp #status predicted
F;1-423/Active site: His (phosphohistidine intermediate) #status predicted
F;92,133,167,177,191,267,322,331/Binding site: carbohydrate (Asn) (covalent) #status pr
F;[illegible]

Query Match 5.1%; Score 193.5; DB 1; Length 423;
Best Local Similarity 23.1%; Pred. No. 2.2e-17;
Matches 98; Conservative 44; Mismatches 180; Indels 104; Gaps 19;

RESULT 2
I16058
hypothetical protein F16D11.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: I16058
R;Fulton, L.
Submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid F16D11.
A;Reference number: S69020
A;Accession: I16058
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-416 <FUL>
A;Cross-reference: EMBL:U40999; NID:g1072175; PIDN:AAA81[illegible]; PID:g1072176; GENP:AAA81[illegible]; GESE:F16
C;Genetics:
A;Gene: CESP:F16D11.1
A;[illegible]
C;Superfamily: mammalian acid phosphatase

Query Match 5.0%; Score 191; DB 2; Length 416;
Best Local Similarity 23.4%; Pred. No. 4.6e-17;
Matches 94; Conservative 43; Mismatches 174; Indels 103; Gaps 19;

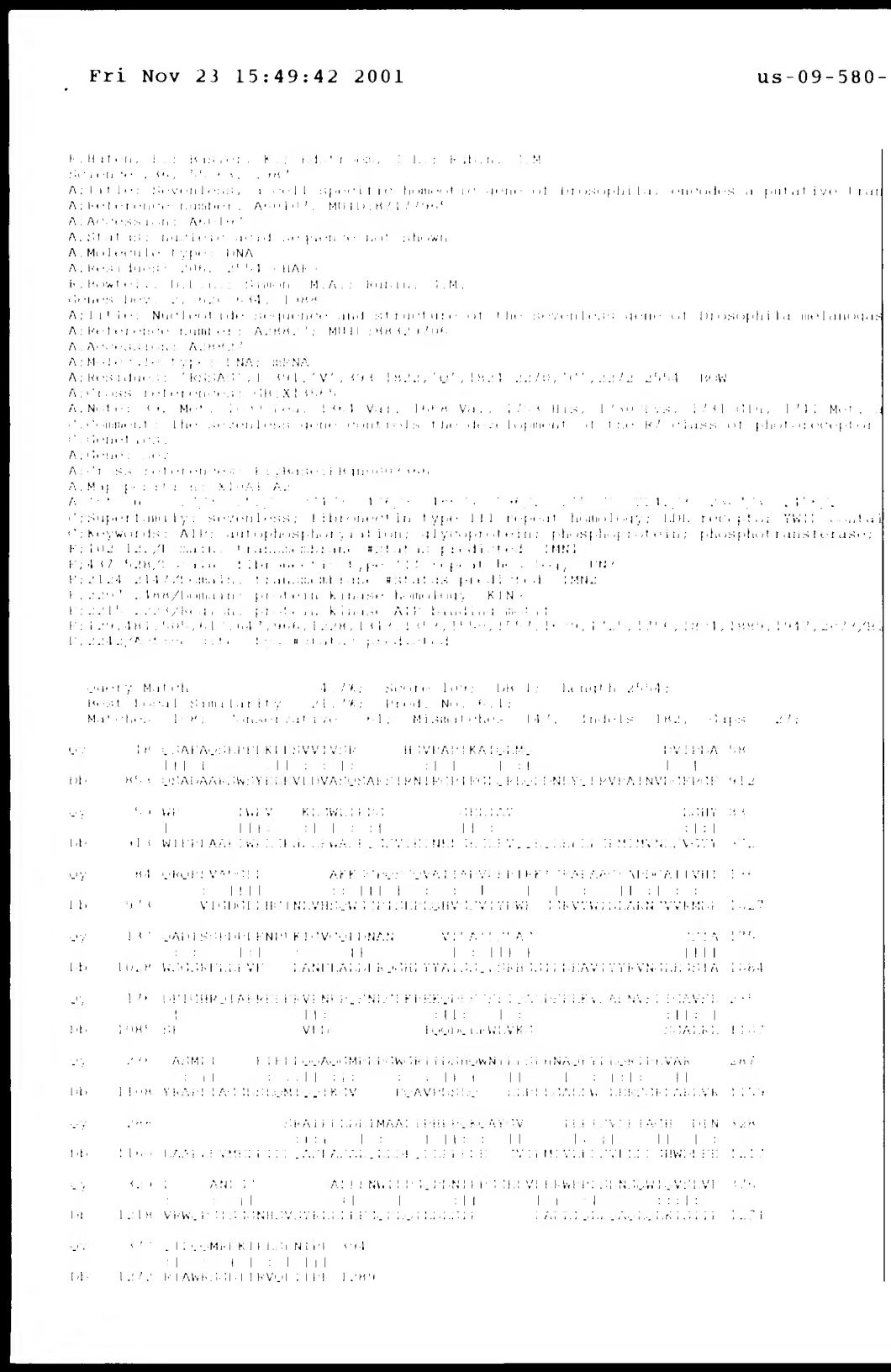


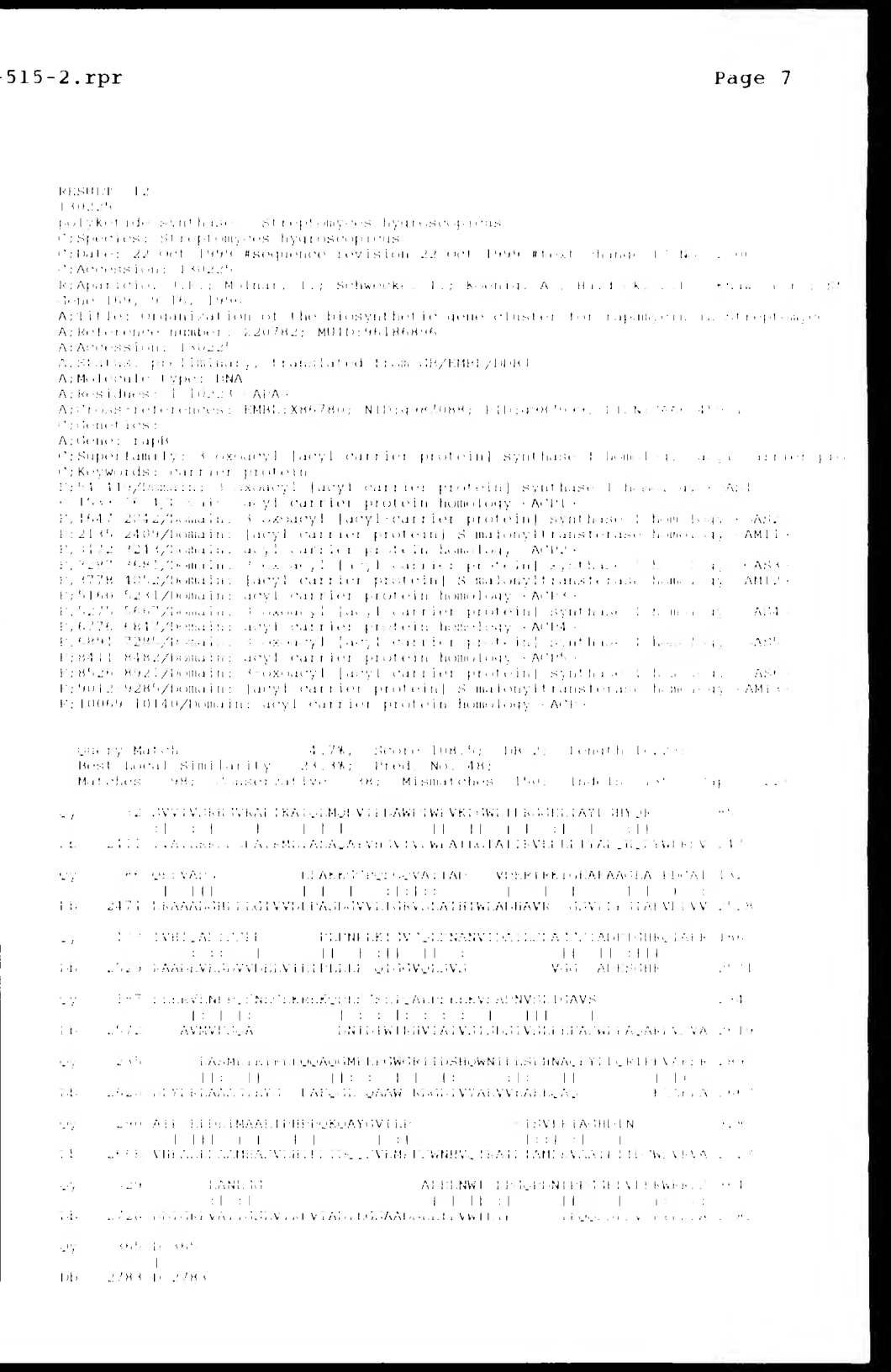
RESULT 6
A44405
acid phosphatase (EC 3.1.3.2) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Jan-1990 #sequence_revision 20-Jan-1990 #text_change 22-Jun-1999
C;Accession: A44405
R;Brunetti, M.; Iuppini, M.; Noguchi, Y.; Kondo, A.; Kato, K.
Biochem. Biophys. Res. Commun. 162, 1044-1051, 1989
A;Title: Isolation and sequencing of a cDNA for rat liver acid phosphatase
A;Reference number: A44405; MUID:89355010
A;Accession: A44405
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <BRM>
A;Cross-references: GB:M27011; NID:g202933; PIDN:AAA41011.1; PID:g202934
C;Superfamily: mammalian acid phosphatase
C;Keywords: glycoprotein; phosphatase; phosphoprotein; phosphoric monoester hydrolase
F;41/Active site: Arg #status predicted
F;142/Active site: Asp (phosphate-histidine intermediate) #status predicted

RESULT 7
S14742
acid phosphatase (EC 3.1.3.2) precursor - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 21-Nov-1993 #sequence_revision 19-Nov-1995 #text_change 31-Mar-2000
C;Accession: S14742
R;Golgi, F.; von Figura, K.; Pohlmann, R.
Biol. Chem. Hoppe-Seyler 372, 301-304, 1991
A;Title: Molecular cloning of the mouse lysosomal acid phosphatase
A;Reference number: S14742; MUID:91282986
A;Accession: S14742
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-421 <GFI>
A;Cross-references: EMBL:X57199; NID:g52870; PIDN:CAA40485.1; PID:g52871
C;Superfamily: mammalian acid phosphatase
C;Keywords: lysosome; phosphoric monoester hydrolase

RESULT 10
JH0610
acid phosphatase (EC 3.1.3.2) ACPP precursor [validated] - human
N;Alternate names: acid phosphatase, prostatic; orthophosphoric monoester phosphohydr
C;Species: Homo sapiens (man)
C;Date: 17-Aug-1992 #sequence_revision 01-Dec-1995 #text_change 09-Jun-2000
C;Accession: JH0610; A39644; A39645; A42419; S71112; S78049; S41761; S41762;
R;Sharief, F.S.; Li, S.S.L.
Biochem. Biophys. Res. Commun. 184, 1468-1476, 1992
A;Title: Structure of human prostatic acid phosphatase gene.
A;Reference number: JH0610; MUID:92272747
A;Accession: JH0610
A;Molecule type: DNA
A;Residues: 1-386 <SHA>
A;Cross-references: GB:M97580; GB:M97581; GB:M97582; GB:M97583; GB:M97584; GB:M97585;
A;Accession: JH0698
A;Molecule type: mRNA
A;Residues: 1-386 <SHA>
A;Cross-references: GB:M97589; NID:g180611; PIDN:AAA60254.1; PID:g180612
R;Van Etten, R.L.; Davidson, K.; Stezia, P.E.; MacArthur, B.; Mueller, G.
J. Biol. Chem. 266, 2413-2419, 1991
A;Title: Covalent structure, disulfide bonding, and identification

```
A;Reference number: A38608; MUID:91175848
A;Accession: A38608
A;Molecule type: mRNA
A;Residues: 1-386 <VAN>
A;Cross-references: GB:M34840; NID:g189021; PIDN:AAA60021.1; PID:g189022
A;Note: part of this sequence, including the amino end of the mature protein, was confir
R;Vihko, P.; Virkkunen, P.; Henttu, P.; Roiko, K.; Soelin, T.; Huhtala, M.L.
FEBS Lett. 236, 275-281, 1988
A;Title: Molecular cloning and sequence analysis of cDNA encoding human prostatic acid p
A;Reference number: S01331; MUID:88312981
A;Accession: S01331
A;Molecule type: mRNA
A;Residues: 1-14,'A',16,'AS',20,'CF',23,'G',25-65,'WIWRIHFA',74-211,'A',213-386 <VIH>
A;Cross-references: EMBL:X52174; NID:g28421; PIDN:CAA36422.1; PID:g28422
A;Note: part of this sequence, including the amino end of the mature protein, was confir
R;Sharief, F.S.; Lee, H.; Leuderman, M.M.; Lundwall, A.; Deaven, L.L.; Lee, C.; Li, S.S.
Biochem. Biophys. Res. Commun. 160, 79-86, 1989
A;Title: Human prostatic acid phosphatase: cDNA cloning, gene mapping and protein sequen
A;Reference number: A32419; MUID:89228674
A;Accession: A32419
A;Molecule type: mRNA
A;Residues: 1-14,'A',16,'AS',20,'CF',23,'G',25-94,'Q',96-115,'R',117-214,'S',216-253,'I
A;Cross-references: GB:M24902; NID:g189018; PIDN:AAA60022.1; PID:g189019
A;Note: the authors translated the codons CAA for residue 95 as Glu, CGT for residue 116
R;Tailor, P.G.; Govindan, M.V.; Patel, P.C.
Nucleic Acids Res. 18, 4928, 1990
A;Title: Nucleotide sequence of human prostatic acid phosphatase determined from a full-
A;Reference number: S11147; MUID:90378494
A;Accession: S11147
A;Status: preliminary; translation not shown
A;Molecule type: mRNA
A;Residues: 1-14,'AFASC',20,'CF',23,'G',25-45,'H',47-65,'KIWIIHFA',74-138,'E',140-156,'R
A;Cross-references: EMBL:X53605; NID:g35683; PIDN:CAA37677.1; PID:g35684
R;Banas, B.; Blaschke, D.; Fittler, F.; Hoerz, W.
submitted to the EMBL Data Library, April 1993
A;Description: Characterization of the promoter of the human prostatic acid phosphatase
A;Reference number: S38863
A;Accession: S38863
A;Molecule type: DNA
A;Residues: 1-40 <BAN>
A;Cross-references: EMBL:X71331; NID:g416730; PIDN:CAA50534.1; PID:g416731
R;Virkkunen, P.H.; Hedberg, P.; Palvimo, J.J.; Birr, E.; Porvari, K.; Vaaraniemi, T.;
submitted to the EMBL Data Library, September 1993
A;Description: Structural organization of human and rat prostate specific acid phosphata
equence in the human gene promoter.
A;Reference number: S41251
A;Accession: S41251
A;Molecule type: DNA
A;Residues: 1-40 <VIR>
A;Cross-references: EMBL:X74961; NID:g397657; PIDN:CAA52910.1; PID:g397658
R;Lee, H.; Chu, T.M.; Li, S.S.L.; Lee, C.
Biochem. J. 277, 759-765, 1991
A;Title: Homodimer and heterodimer subunits of human prostate acid phosphatase.
A;Reference number: S17042; MUID:91339999
A;Accession: S17042
A;Status: preliminary
A;Molecule type: protein
A;Residues: 33-49 <LEE>
R;Banas, B.; Blaschke, D.; Fittler, F.; Hoerz, W.
Biochim. Biophys. Acta 1217, 188-194, 1994
A;Title: Analysis of the promoter of the human prostatic acid phosphatase gene.
A;Reference number: S42730; MUID:94153595
A;Accession: S42730
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-40 <BA2>
A;Cross-references: GB:X71331; NID:g416730; PIDN:CAA50534.1; PID:g416731
R;Sharief, F.S.; Li, S.S.
Biochem. Mol. Biol. Int. 33, 561-565, 1994
A;Title: Nucleotide sequence of human prostatic acid phosphatase ACPP gene.
A;Reference number: I37175; MUID:95038536
A;Accession: I37175
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-386 <RES>
A;Cross-references: EMBL:U07807; NID:g[illegible]; PIDN:AAB60640.1; PID:g[illegible]
C;Comment: This protein is synthesized under androgen regulation by epithelial cells
C;Genetics:
A;Gene: GDB:ACPP
A;Cross-references: GDB:119644; OMIM:171790
A;Map position: 3q21-3q23.2
A;Introns: [illegible]
C;Function:
A;Description: catalyzes the hydrolysis of a wide range of phosphate esters
C;Superfamily: mammalian acid phosphatase
C;Keywords: glycoprotein; phosphohistidine; phosphoprotein; phosphoric monoester hydr
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-386/Product: acid phosphatase ACPP #status experimental <MAT>
F;43,85/Active site: Arg #status predicted
F;44/Active site: His (phosphohistidine intermediate) #status predicted
F;94,220,333/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;161-372,215-313,347-351/Disulfide bonds: #status experimental

  Query Match             5.1%;  Score 117;  DB 1;  Length 386;
  Best Local Similarity   19.8%;  Pred. No. 0.096;
  Matches   87;  Conservative   6[illegible];  Mismatches  178;  Indels  11[illegible];  Gaps  15;

Qy      [illegible] AII[illegible]QAIAQ[illegible]LLKLLSVVIVG[illegible]VKA[illegible]TKA[illegible]AWL [illegible]
            :: : || ||      :|  |:     :|: | :| ||| |:|             | .|.
Db      [illegible] [illegible]LLFTWLLRSVLAK      LLKFVILVFRHGDRSPI        -    [illegible]

Qy      [illegible]        [illegible]ADVI 112
               ||    | ||  | |    || : ::         :|   :|      |   |
Db      57 FIELKWE[illegible]YTEKET        LKE[illegible]V 109

Qy      113 ERTEK[illegible]EAIAACLAP[illegible]AIIVHIQALI[illegible]LFNPLKI[illegible]NANVIDAII[illegible] 172
            :||    :        | | |    :::        :|   |: |:        |
Db      110 [illegible]NLAALFPFEGVGIW        NPILLWQPIPVHIVPLSE[illegible] 150

Qy      173 [illegible] 232
                              :::|   |   |  |  | : :: :       ::      |      | :: |
Db      151                  D[illegible]LYLPFRN[illegible]PRE[illegible]LKSEE[illegible]KRLPPYK[illegible]IATL[illegible]K 193

Qy      233 V[illegible]           LASN[illegible]QQAQ[illegible]WNI[illegible]ENA[illegible]YLLQKIF 283
            :                :  |  :  :      :         | |            :     |
Db      194 LSGLHGQDLFGIWSKVYDPLYCESVHNFTLPSWATEDTMTKLRELSELSLLSLYGIHKQK 253

Qy      284 [illegible] 342
            | :|  :     |:: |:   :           |:|    :           ::   : |||  :: |    |||:  |   |
Db      254 EKSR[illegible]IVLVNHI[illegible]NHH------KKAI[illegible]YKKLIMYSAHDTTVS[illegible]QMALDVYN[illegible] [illegible]

Qy      [illegible] [illegible]EKTPLSENIPPIPVKIIIA 402
            |   |   :       || |:   :               : : ::                       .
Db      308 P  PYASCHLTELYFEKGEYF          VEMYYRN               ETQHEPYPLMLP 345

Qy      403 G[illegible]HEKNAQ[illegible]LAGPIQIV 422
            :|             | :  | ::|
Db      346 GCSP      SCPLERFAELV 366


RESULT 11
[illegible]
Kinase-related protein sevenless - fruit fly (Drosophila melanogaster)
N;Contains: protein-tyrosine kinase (EC 2.7.1.112)
C;Species: Drosophila melanogaster
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change [illegible]
C;Accession: A26912; A60[illegible]; A26827
R;Basler, K.; Hafen, E.
Cell 54, 299-311, 1988
A;Title: Control of photoreceptor cell fate by the sevenless protein requires a functi
A;Reference number: A26912; MUID:88282538
A;Accession: A26912
A;Molecule type: DNA
A;Residues: 1-2554 <BAS>
A;Cross-references: GB:[illegible]; NID:g157413; PIDN:AAA[illegible]
```





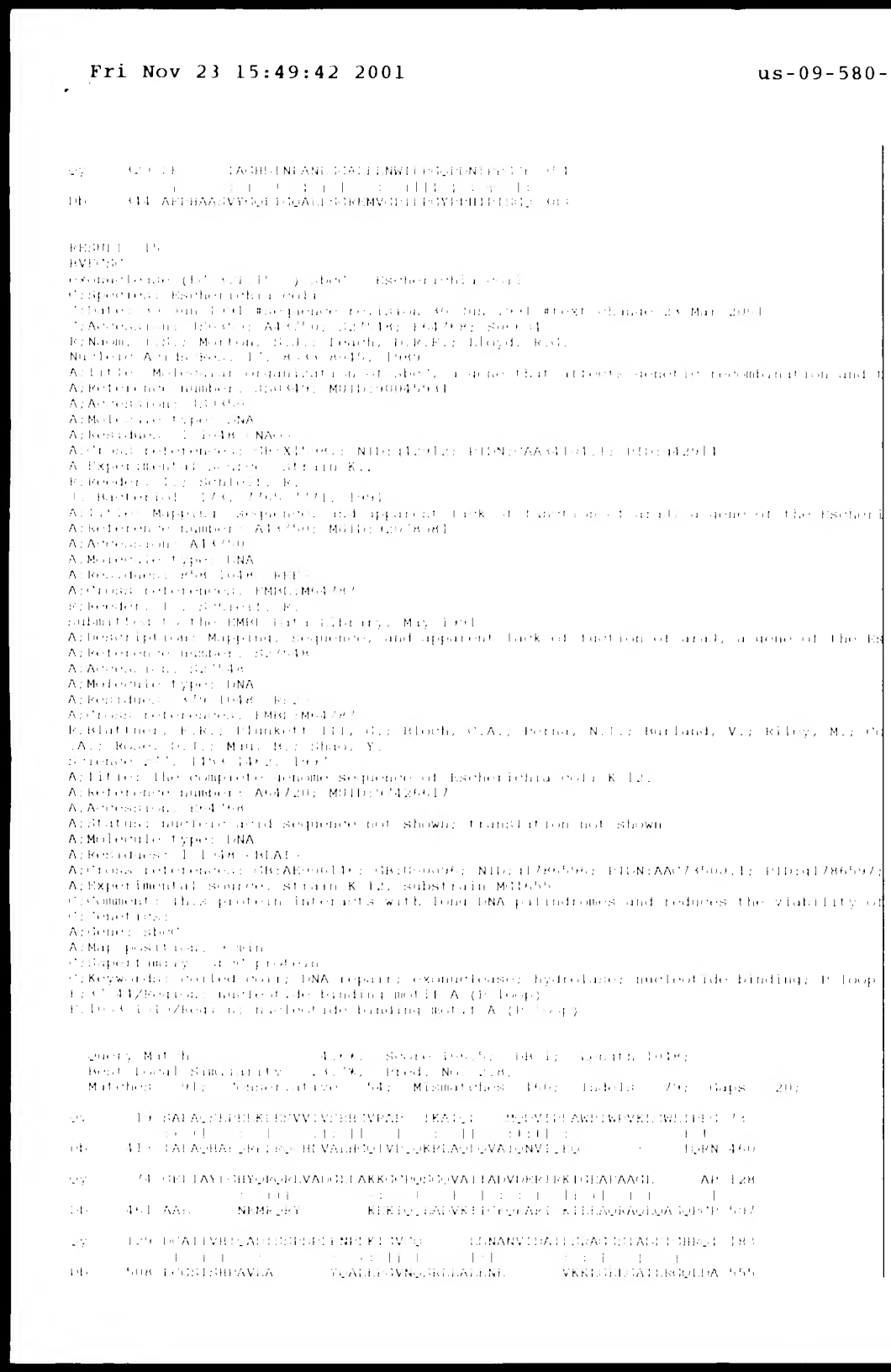
RESULT 1
A54429
paired box transcription factor PAX8 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change 16-Jul-1999
C;Accession: A54429
R;Plachov, D.; Chowdhury, K.; Walther, C.; Simon, D.; Guenet, J.L.; Gruss, P.
Development 110, 643-651, 1990
A;Title: Pax8, a murine paired box gene expressed in the developing excretory system and
A;Reference number: A54429; MUID:91215711
A;Accession: A54429
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-457 <PLA
A;Cross-references: GB:X57467; NID:g56566; PIDN:CAA40761.1; PID:g56564
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIN:M31850, NCBIP:M31850)
C;Superfamily: paired box transcription factor Pax 8; paired box homology
F;1-130/Domain: paired box homology <PBH

Query Match 4.7%; Score 107; DB 2; Length 457;
Best Local Similarity 21.1%; Pred. No. 0.73;
Matches 84; Conservative 59; Mismatches 151; Indels 105; Gaps 19;

RESULT 2
A54429
paired box transcription factor PAX8, splice form a - human
N;Contains: paired box transcription factor PAX8, splice form b
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1995 #sequence_revision 30-Jun-1995 #text_change 16-Jul-1999
C;Accession: A54429; A27967; S78492; I67475; A45525
R;Kozmik, Z.; Kurzbauer, R.; Dorfler, P.; Busslinger, M.
Mol. Cell. Biol. 13, 6024-6035, 1993
A;Title: Alternative splicing of Pax-8 gene transcripts is developmentally regulated and
A;Reference number: A54429; MUID:94017277
A;Accession: A54429
A;Molecule type: mRNA; DNA
A;Residues: 1-450 <KOZ>
A;Cross-references: GB:L19686; NID:g397320; PIDN:AAA36570.1; PID:g397321
R;Plachov, D.; Chowdhury, K.; Mundlos, S.; Winterpacht, A.; Zabel, B.; Fidler, A.; Gruss
Development 116, 611-623, 1992

A;Title: PAX8, a human paired box gene: isolation and expression in developing thyroid
A;Reference number: A49126; MUID:93170167
A;Accession: S78492
A;Molecule type: mRNA
A;Residues: 1-417,'G',419-450 <PHL>
A;Cross-references: EMBL:X69699; GB:S78490; NID:g48425
A;Experimental source: kidney
A;Note: this CDS is not annotated in GenBank entry HSPAX8A, release 33
R;Plachov, D.
Submitted to the EMBL Data Library, February 1993
A;Reference number: S78492
A;Accession: S78492
A;Molecule type: mRNA
A;Residues: 1-304,'L',306-321,'G',323-428,'L',430-417,'G',419-450 <HIA>
A;Cross-references: EMBL:X69699
R;Plachov, D.; Beier, K.; Bleckmann, H.; Simmler, M.C.; Taviaux, S.; Attie, T.;
EMBO J. 9, 3169-3178, 1990
A;Title: Distinct functional properties of three human paired box proteins, PAX8, PAX
A;Reference number: 15340; MUID:90255300
A;Accession: I67474
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-259,'M',261-417,'G',419-450 <PHL>
A;Cross-references: GB:S27905; NID:g297792; PIDN:AAB34174.1; PID:g297793
A;Note: this is an incorrect construction of splice form d
A;Accession: I67475
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-256,427-450 <PHL2>
A;Cross-references: GB:S27906; NID:g297794; PIDN:AAB34176.1; PID:g297795
A;Note: this is an incorrect construction of splice form e
C;Comment: Four splice forms: a, d, and e, see PIR:I15460.
C;Genetics:
A;Gene: GB:PAX8
A;Cross-references: GDB:136447; OMIM:167415
A;Map position: 2q12-2q14
A;Introns: 52/3 <2> 84/1 <3> 126/3 <4> 262/3 <5> 315/2 <6> 369/3 <7> 419/1 <8>
A;Note: 45 nt 5' untranslated; 1-66=poor splice boundary at 66/2 probable frameshifted d
o
C;Superfamily: paired box transcription factor Pax 8; paired box homology
C;Keywords: alternative splicing; DNA binding; transcription regulation
F;1-45%/Disorder: paired box transcription factor PAX8, splice form a #status predicted
F;1-299,368-450/Product: paired box transcription factor PAX8, splice form b #status
F;1-130/Domain: paired box homology <PBH>

Query Match 4.6%; Score 107; DB 2; Length 450;
Best Local Similarity 22.0%; Pred. No. 0.77;
Matches 88; Conservative 59; Mismatches 139; Indels 114; Gaps 21;



RESULT 15
sbcC
exonuclease (EC 3.1.-.-) sbcC - Escherichia coli
C;Species: Escherichia coli
C;Date: 30-Jun-1994 #sequence_revision 30-Jun-1994 #text_change 25-Mar-2001
C;Accession: D64754; A41750; S27198; I64798; S00134
R;Naom, I.S.; Morton, S.J.; Leach, D.R.F.; Lloyd, R.G.
Nucleic Acids Res. 17, 6549-6571, 1989
A;Title: Molecular organization of sbcC, a gene that affects genetic recombination and t
A;Reference number: A30484; MUID:90045041
A;Accession: B31850
A;Molecule type: DNA
A;Residues: 1-1618 <NAV>
A;Cross-reference: GB:X15996.1; NID:g41201; PIDN:CAA33181.1; PID:g41211
A;Experimental source: strain K..
R;Kneale, G.; Sedgwick, B.
J. Bacteriol. 174, 7565-7571, 1992
A;Title: Mapping sequences and apparent lack of function of sbcB, a gene of the Escheri
A;Reference number: A41750; MUID:92278081
A;Accession: A41750
A;Molecule type: DNA
A;Residues: 858-2040 <KEE>
A;Cross-reference: EMBL:M64787
R;Kneale, G.; Sedgwick, B.
Submitted to the EMBL Data Library, May 1991
A;Description: Mapping, sequence, and apparent lack of function of sbcB, a gene of the Es
A;Reference number: S27548
A;Accession: S27548
A;Molecule type: DNA
A;Residues: 858-2040 <KEE>
A;Cross-reference: EMBL:M64787
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: B64756
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1618 <BLA>
A;Cross-reference: GB:AE000410; GB:U00006; NID:g1786596; PIDN:AAC73500.1; PID:g1786597
A;Experimental source: strain K-12, substrain MG1655
C;Comment: sbcC protein interacts with four DNA palindromes and reduces the viability of
C;Genetics:
A;Gene: sbcC
A;Map position: 0 min
C;Superfamily: sbcC protein
C;Keywords: coiled coil; DNA repair; exonuclease; hydrolase; nucleotide binding; P loop
F;1-44/Region: nucleotide binding motif A (P loop)
F;1-44/Binding site: nucleotide binding motif A (P loop)

Query Match 4.6%; Score 1055.5; DB 1; Length 1618;
Best Local Similarity 43.7%; Pred. No. 2.6e-48;
Matches 911; Conservative 564; Mismatches 1106; Indels 79; Gaps 20;